Manuel Maria Reis de Orey
Licenciado em Química Aplicada

Previsão de propriedades elétricas de células solares sensibilizadas por corantes aplicando QSPR

Dissertação para obtenção do Grau de Mestre em
Química Bioorgânica

Orientador: João Aires de Sousa, Professor Auxiliar com Agregação, DQ FCT/UNL

Co-orientador: João Carlos Lima, Professor Associado com Agregação, DQ FCT/UNL

Júri:

Presidente: Professora Doutora Paula Cristina de Sério Branco

Arguente: Professor Doutor André Osório Falcão

Vogal: Professor Doutor João Montargil Aires de Sousa

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Março, 2020

Previsão de propriedades elétricas de células solares sensibilizadas por corantes aplicando QSPR

Agradecimentos

Primeiramente queria agradecer ao professor João Aires de Sousa pela orientação, apoio na elaboração deste trabalho e pelo espaço de trabalho disponibilizado, ao professor João Carlos Lima, pela ajuda na elaboração deste trabalho, a professora Paula Branco, pelo apoio neste trabalho e durante o mestrado, a investigadora Florbela Pereira pela sua assistência e a FCT/UNL. Também queria agradecer aos meus pais, irmãos e amigos por me ajudarem durante esta fase na minha vida.

Resumo

Atualmente a transição de fontes de energia não-renováveis para fontes de energia renováveis assumiu maior relevância e procura. As células solares sensibilizadas por corantes são um tipo de células solares com grande potencial para a geração de energia. O objetivo deste trabalho consiste em construir um modelo QSPR para a previsão dos parâmetros das células associados aos corantes usados. Para alcançar esse objetivo foram usados algoritmos de Random Forest e redes neuronais. Foram testados vários conjuntos e subconjuntos. Inicialmente utilizou-se um conjunto total de 882 células, sendo depois separado em subconjuntos de forma a agrupar corantes e células com semelhanças. Foram criados três subconjuntos, que foram também processados de forma independente: o primeiro com corantes onde existe uma unidade de cumarina, o segundo com as células onde existe a presença de t-butil piridina no eletrólito e o último com as células onde existe t-butil piridina e também uma camada de scattering. O melhor modelo conseguido tem um RMSE de 1,345 na previsão da percentagem da eficiência de conversão (PCE) para um conjunto independente de teste. Na previsão do FF e PCE foram identificadas dificuldades não associadas ao corante, mas sim derivadas da construção da célula.

Palavras-chave: QSPR, DSSC, células sensibilizadas por corantes, Random Forest, rede neuronal

Abstract

At a time when there is a need to transition from non-renewable energy to renewable energy, one of the existing solutions are solar cells. Dye-sensitized solar cells are a type of solar cell with immense potential in the generation of energy. The objective of this paper consisted in building a QSPR model to predict parameters of the solar cells corresponding to the used dyes. To achieve that objective, a Random Forest algorithm and neural network methods were employed. Various sets and subsets were tested. Initially a set of the total 882 dyes was used, which was then separated into subsets to group together dyes with similar characteristics. Three subsets were also made, which were processed independently: the first includes all dyes with the coumarin unit, the second where there is the presence of t-butyl pyridine in the electrolyte, and the last subset also included dyes with t-butyl pyridine but also with a scattering layer in the cell. The best model achieved has a RMSE of 1.345 in the prediction of the percentage efficiency (PCE). It was concluded that there are problems with the prediction of the FF and PCE, unrelated to the dyes, linked with the construction of the cell.

Keywords: QSPR, DSSC, dye-sensitized solar cell, Random Forest, neural network

Índice de conteúdo

Índice de figuras

Índice de Tabelas

Índice de Equações

1. Introdução

1.1 Contextualização

1.1.1 História

As civilizações ao longo dos tempos sempre tiveram necessidades energéticas, quer para a produção alimentar, quer para o aquecimento e iluminação. A fonte de combustível variou entre o que havia mais por perto e se encontrava disponível.[1] Com o passar do tempo, várias foram as mudanças nas fontes de energia o que dependeu de vários fatores: disponibilidade, avanços tecnológicos, impacto na saúde ou ambiente e custo. Essas mudanças sempre foram a um nível local, regional ou até individual.

A população humana está em crescimento, aumentando dois mil milhões numa só geração, colocando em causa a sustentabilidade das atuais fontes de energia.[2] Outro fator importante é o efeito que as atuais fontes de energia têm, tanto em termos de saúde,[3] como em termos ambientais, sendo o aquecimento global um problema que pode ter consequências graves.[4]

1.1.2 Fontes alternativas de produção de energia

Existem várias alternativas,[5] como por exemplo: biomassa, hidroelétrica, vento, fotovoltaica, fototérmica, entre outras. Alguns exemplos são as células solares ou a captação e distribuição de calor para aquecimento doméstico.[6]

Comparando com as outras alternativas, em 2011 um relatório produzido para o parlamento britânico[7] concluiu que painéis fotovoltaicos de silício têm das maiores pegadas de carbono (fig. 1.1). No entanto também refere que, dadas as melhorias na produção, essa pegada está a diminuir, com novas alternativas a baixar para a parte inferior do intervalo.

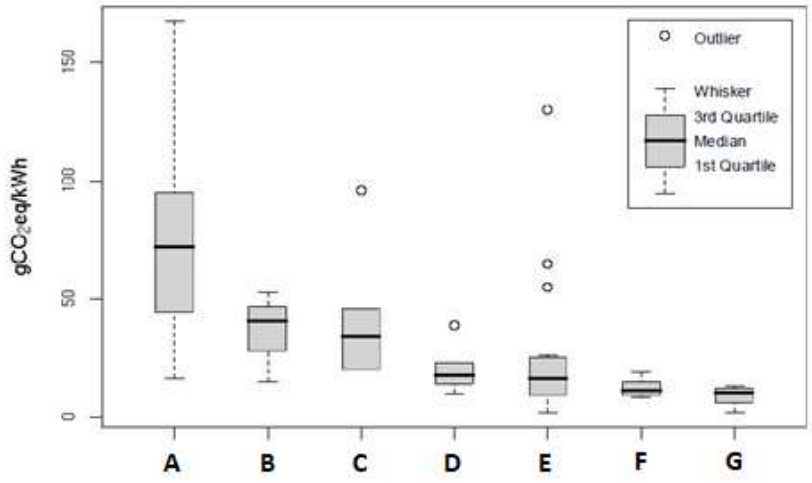

Figura 1.1 Pegada de carbono de várias formas de produção de energia[7] A) Solar fotovoltaica B) Geotérmica C) Eólica de pequena/média escala D) Marinha E) Nuclear F) Eólica de grande escala G) Hidroelétrica de rios

1.1.3 **Energia solar**

Um aspeto positivo da exploração da energia solar é o seu potencial. Existem 6 zonas no mundo, que se exploradas, poderiam fornecer toda a energia necessária para o mundo, mesmo com eficiências de conversão de somente 8%.[6] Usando os Estados Unidos da América como exemplo, em 2017 o total de energia elétrica produzida foi 4034 TWh,[8] com o potencial fotovoltaico sendo estimado em 208613 TWh.[9]

Das centrais elétricas que usam células solares, 93% dos painéis são à base de silício.[10] Essas células têm um potencial máximo teórico de 31%,[11] tendo sido atingida uma eficiência máxima registada de 26,7%[12] em 2018. No entanto, é necessário silício altamente puro, que requer alto vácuo e temperaturas elevadas na sua manufatura.

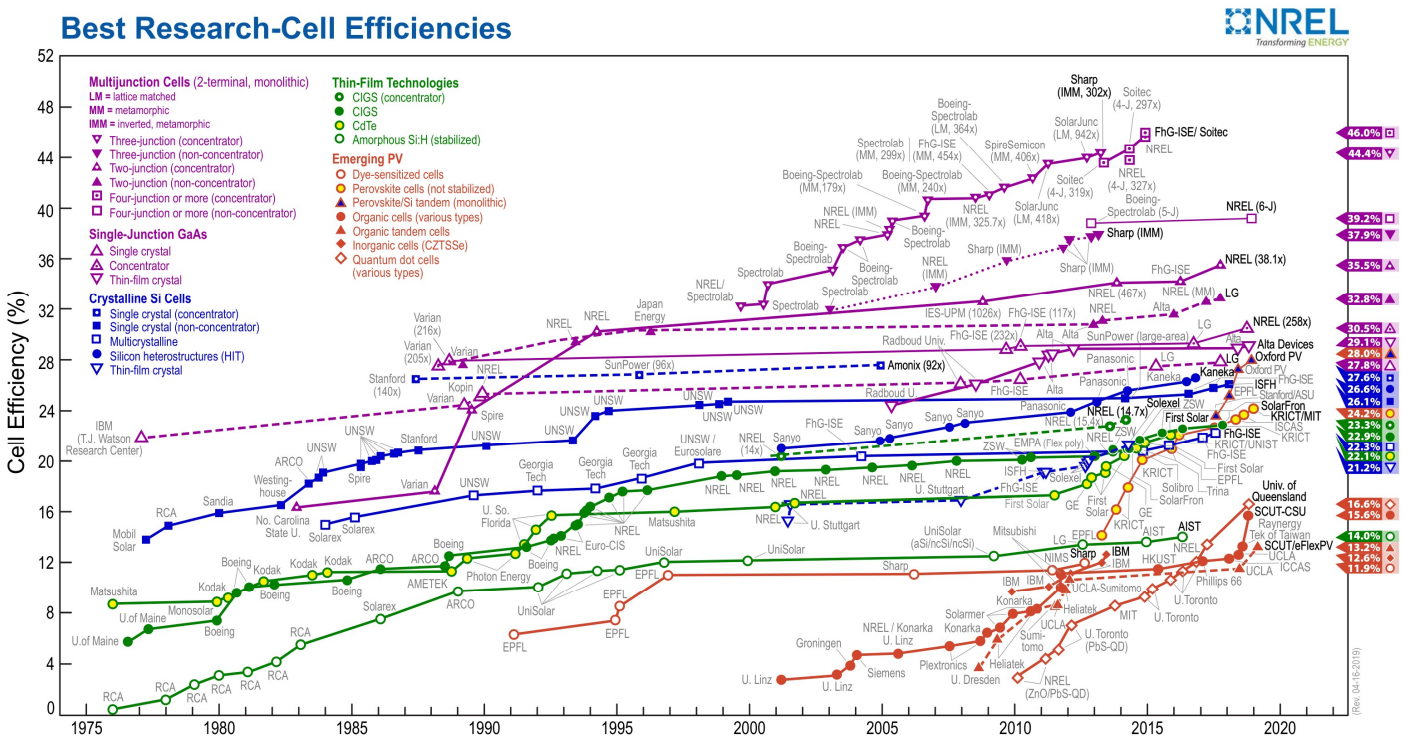

Figura 1.2 Desenvolvimento de células solares[13]

Na figura 1.2 está demostrado o desenvolvimento de vários tipos de células solares ao longo dos anos.

1.2 **Células solares sensibilizadas por corantes**

1.2.1 **Descrição**

Desde a publicação de O'Regan e Grätzel[14] em 1991 tem havido grande interesse nas células solares sensibilizadas por corantes (*Dye-sensitized solar cells* ou DSSCs).[15–19] A utilização de DSSCs tem alguns problemas, mas também tem benefícios quando comparada com a de outras formas de células solares.

Uma desvantagem é a utilização de solventes orgânicos, que têm o potencial de permear pelo plástico o que traz um potencial perigo ambiental.[20] No entanto já existe pesquisa e trabalho em células sem solventes, resolvendo este problema.[21–23]

Com o baixar dos custos de produção,[24] e a sua flexibilidade de utilização demostrada pela a possibilidade de imprimir células,[19] existe um grande potencial para DSSCs, existindo no entanto a necessidade de aumentar a eficiência para tornar a tecnologia mais competitiva.

1.2.2 **Funcionamento de DSSCs**

O funcionamento de DSSCs é muito análogo ao funcionamento da clorofila nas plantas. O corante é adsorvido à superfície de um semicondutor, e quando absorve um fotão (fotoexcitação), passa para o estado excitado. Seguidamente, o eletrão é transferido para o semicondutor (injeção), passando deste para o circuito externo, onde é debitado no contra elétrodo, produzindo trabalho elétrico. Depois regressa ao corante via um eletrólito (regeneração) transportador de eletrões, que é reduzido no contra elétrodo e reoxidado no ato de regenerar o corante, reduzindo a sua forma oxidada e voltando ao estado fundamental inicial (fig. 1.3).

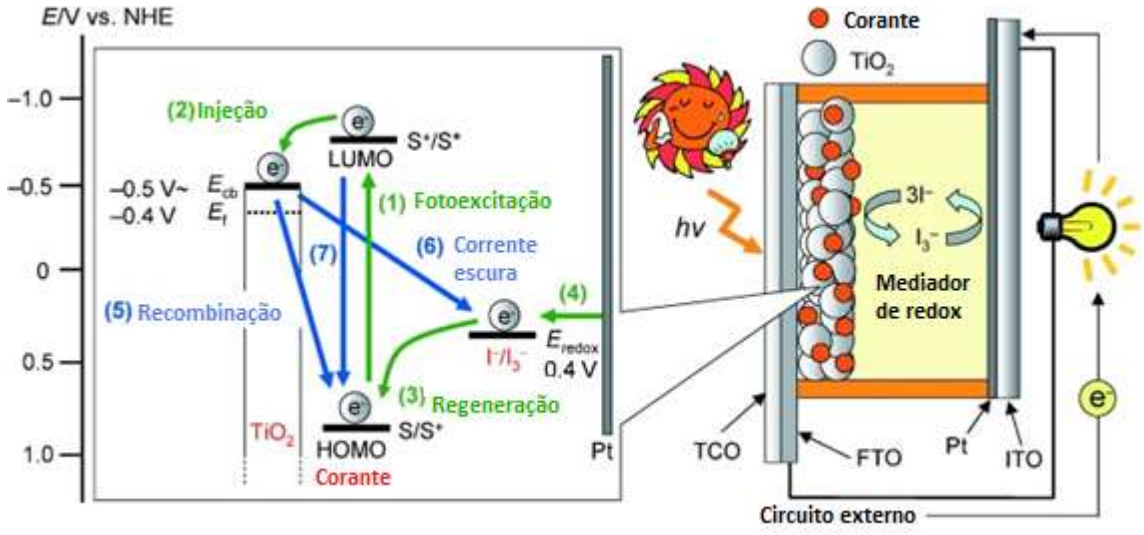

Figura 1.3 Ilustração do funcionamento de um DSSC[16]

Atualmente existe muita pesquisa no modelo de Doador-Ponte conjugada-Aceitador (fig. 1.4), em que existe um grupo que doa o eletrão, uma ponte para estender o máximo de absorção do corante para o visível e um aceitante que serve simultaneamente como o ponto de ancoragem ao semicondutor e permite a injeção do eletrão neste.

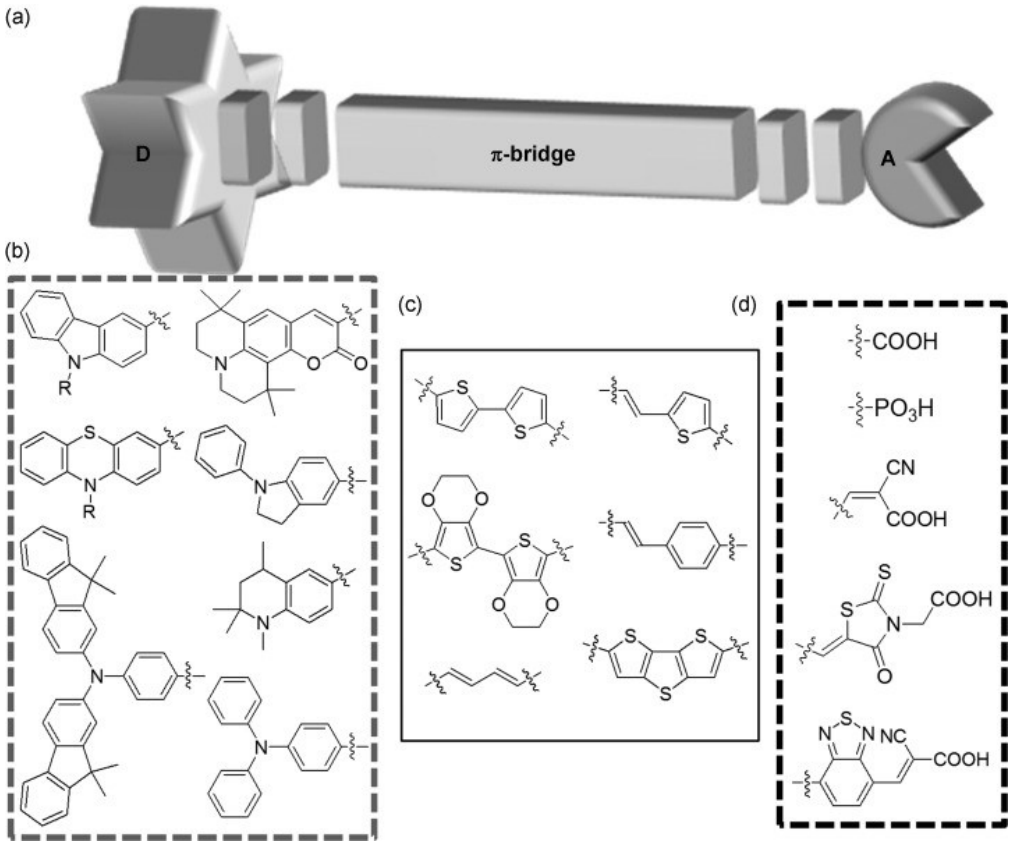

Figura 1.4 a) Esquemática do modelo Doador-Ponte conjugada-Aceitante para DSSCs[15] b) doadores c) pontes d) aceitantes/grupos de ancoragem

Existem várias famílias de moléculas utilizadas em DSSCs[25] (fig. 1.4), como por exemplo a trifenilamina ou a cumarina, como o doador, grupos de tiofeno ou fenil, como a ponte conjugada, e o grupo mais comum de aceitante é o ácido ciano acrílico, que também funciona como grupo de ancoragem.[15]

A eficiência de conversão fotovoltaica, PCE(%), de uma DSSC é dada pela equação (1). I_0 é a potência por cm^2 da luz irradiada, normalmente em condição padrão de 100 $mWcm^{-2}$ (AM 1.5).

$$PCE\ (\%)\ =\ \frac{J_{SC}\ (mA\ cm^{-2}) \times V_{OC}\ (V) \times ff}{I_0\ (mW\ cm^{-2})} \qquad (1)$$

J_{SC} é a corrente máxima produzida pela célula em condições de curto-circuito quando a célula é irradiada. É uma medida da interação do corante com o semicondutor (que pode facilitar ou dificultar a injeção de eletrões) e também do coeficiente de absorção molar do corante. Um J_{SC} alto está associado a um corante com uma gama grande de luz visível absorvida, alta eficiência de injeção do eletrão no semicondutor e uma eficiente redução do corante via eletrólito.[16] O J_{SC} pode ser calculado com base na equação (2) com a representação gráfica do IPCE (*Incident-Photon-to-Current Conversion Efficiency*) onde I_S é o fluxo de fotões ao comprimento de onda (λ) em condições AM 1.5 (100 $mWcm^{-2}$) e *e* é a carga do eletrão.

$$J_{SC} = e \int IPCE(\lambda) I_S(\lambda) d(\lambda) \qquad (2)$$

O V_{OC} é a diferença de potencial máxima obtida quando a célula é irradiada num circuito aberto, evitando a passagem de eletrões. Matematicamente é dada pela equação (3).

$$V_{OC} = \frac{E_{cb}}{e} + \frac{k_B T}{e} ln\left(\frac{n}{N_{cb}}\right) - E_{redox} \quad (3)$$

Na equação 3, k_B é a constante de Boltzmann, E_{cb} é a energia da banda condutora do semicondutor, e é a carga do eletrão, T é a temperatura absoluta, n é o número de eletrões no semicondutor, N_{cb} é a densidade eletrónica dos estados e E_{redox} é o potencial de redução do eletrólito. O máximo valor de V_{OC} possível é dado pela diferença entre a E_{cb} e o E_{redox}, e como tal o corante não seria de esperar ter influência sobre esse valor.[16] No entanto fatores como a recombinação e a corrente residual obtida na ausência de irradiação (fig. 1.3) têm impacto no V_{OC}, podendo ser minimizados com o desenho do corante.[15]

O fator de preenchimento (*fil factor*, FF) é a razão entre o potencial máximo da célula (V_{mp} x J_{mp}) e o produto do V_{OC} x J_{SC} [equação (4)]. O FF pode ser determinado pela curva caraterística corrente-tensão da DSSC[16] (fig. 1.5).

$$FF = \frac{V_{mp} \times J_{mp}}{V_{OC} \times J_{SC}} \quad (4)$$

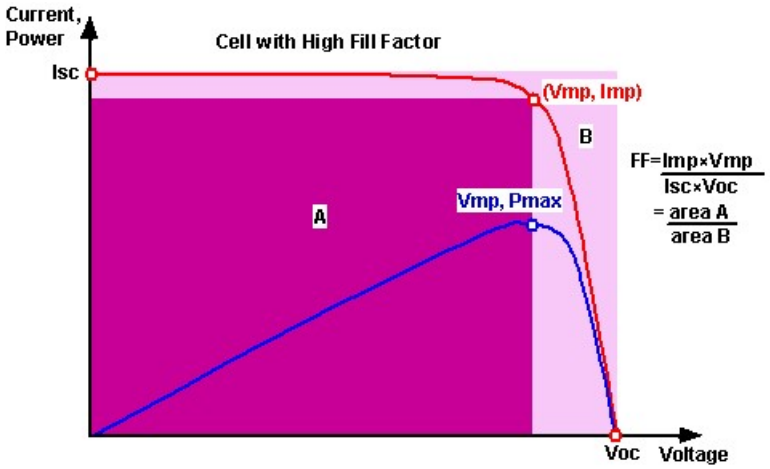

Figura 1.5 Representação da curva corrente-tensão e cálculo do PCE[26]

A interligação entre os diferentes parâmetros que definem a eficiência da célula torna difícil o isolamento de características desejáveis durante o desenho de um corante para DSSC. Um dos objetivos desta tese é a utilização de algoritmos de inteligência artificial para encontrar correlações entre descritores moleculares de corantes e parâmetros de eficiência de DSSCs que permitam abrir novas perspetivas no futuro desenho de corantes para DSSCs.

1.3 **Técnicas de aprendizagem automática**

1.3.1 **Uma perspetiva histórica**

As raízes dos métodos de correlação entre estrutura e atividade (*quantitative structure-activity relationship*, QSAR) podem ser traçadas a 1863, ano em que Cros apresenta a sua tese, em que faz a ligação entre a toxicidade de álcoois amílicos e a suas solubilidades.[27] Estudos QSAR modernos tiveram o início com os trabalhos desenvolvidos por Hansch *et al.* na década de 1960.[28,29]

A utilização de métodos de aprendizagem automática tem vindo a aumentar nos últimos anos, já sendo uma ferramenta essencial na indústria farmacêutica, na forma de QSAR para a descoberta e otimização de fármacos ativos, ou para a eliminação de candidatos potencialmente tóxicos antes que haja muito investimento neles, poupando assim recursos.[30]

O termo *quantitative structure-property relationship* (QSPR) é usado para modelos que preveem outras propriedades de moléculas, que não a sua atividade – como é o caso dos parâmetros associados às células solares.

Recentemente foram publicados trabalhos utilizando métodos para prever propriedades de corantes para DSSCs que utilizam regressão multilinear[31] (MLR) e máquina de vetores de suporte[32] (*support vector machine,* SVM). Nos dois casos é feito um passo inicial de pré-seleção de dados, que pode alterar as conclusões ou a capacidade de previsão do modelo. Ambos os modelos utilizam cálculos de química quântica, aumentando assim o tempo necessário para preparar e aplicar o modelo. A pré-seleção de variáveis pode ser evitada através da utilização do modelo *Random Forest* que pode ainda facilitar a interpretação de resultados e fornecer nova informação sobre quais as propriedades mais importantes a ter em conta quando se está a desenvolver um novo corante.

1.3.2 **O modelo Random Forest**

Desde a sua introdução por Breiman em 2001,[33] *Random Forests* (RF) têm tido aplicações variadas. Por exemplo, no laboratório onde este trabalho foi realizado, têm sido usadas para propriedades relacionadas como a previsão de energias HOMO-LUMO de moléculas[34] ou momento dipolar.[35] Quando comparado com outros métodos, RF apresenta em geral resultados iguais ou melhores,[36] sendo que a sua utilização é bastante simples. O modelo RF necessita tipicamente de duas escolhas do operador,[37] o número de árvores a crescer e o m_{try}, o número de descritores a testar em cada nodo.

Uma RF é um conjunto de árvores de decisão, a previsão final é dada pela votação, em caso de classificação, ou média, em caso de regressão, do valor de saída associado a cada árvore (fig. 1.6).

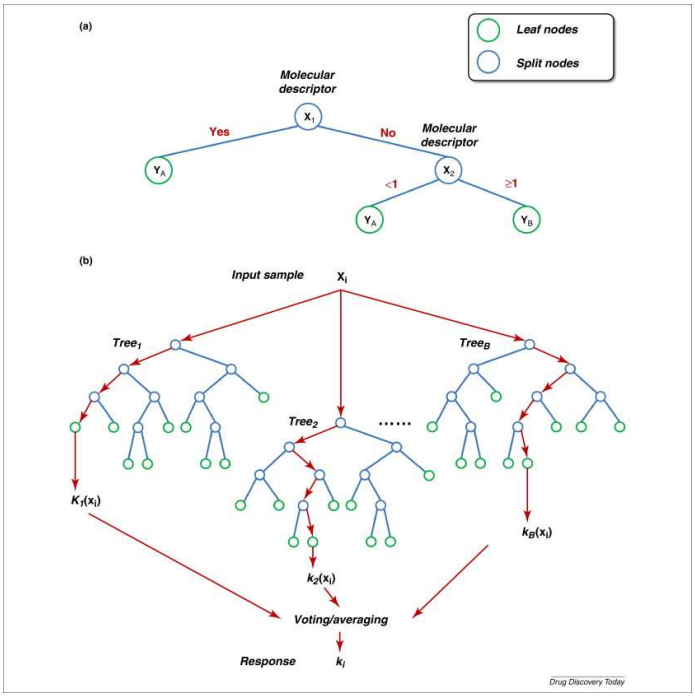

Figura 1.6 Representação de uma Random Forest[38] onde a) como a separação e feita b) como o valor final de output é calculado

Cada árvore individual é construída com um subconjunto aleatório do conjunto de treino, e os dados que ficam de fora compõem o subconjunto chamado *out-of-bag* (OOB). Também para a construção de cada nodo é disponibilizado apenas um número de descritores (m_{try}) selecionados aleatoriamente. O OOB é usado para validar cada arvore.[36,37] Paralelamente, se o conjunto total for grande o suficiente, é possível criar um grupo de teste externo, também para efeitos de validação.

O processo RF consegue diferenciar entre os descritores que têm mais poder discriminativo e os que não têm, sendo que no final do treino é criada uma lista que mede a importância dos descritores. Este processo é capaz de ter como entrada um grande número de descritores, não sendo necessária a pré-seleção de descritores.[36]

A importância dum descritor é dada de duas formas. A primeira pela variação do erro do OOB quando um descritor é permutado. A segunda forma é o cálculo da soma residual dos quadrados (*residual sum of squares*, RSS) onde um descritor é medido com a seguinte equação:

$$RSS = \sum_{i=1}^{N} [y_i - f(x_i)]^2 \quad (5)$$

onde y_i é o valor experimental e x_i é o valor modelado para os elementos *i* que o nodo separou. A diferença entre os valores de RSS antes e depois da separação dos dados com base no descritor é uma medida do poder de diferenciação do descritor.

1.3.3 **Redes Neuronais**

Redes Neuronais ou *artificial neural networks* (ANNs) são um método de aprendizagem automática utilizado nas mais variadas áreas. Relacionado com este trabalho, existem aplicações para prever o tamanho de um sistema fotovoltaico,[39,40] e o máximo de absorção para corantes de DSSC.[41]

As bases matemáticas para redes neuronais foram desenvolvidas por McCulloch e Pitts em 1943.[42] As ANN funcionam de forma simples, havendo uma camada de entrada e uma camada de saída, sendo as camadas intermédias denominadas camadas escondidas (*hidden layers*). As setas entre neurónios denotam o peso aplicado a cada valor que é transportado entre esses neurónios. Na figura 1.7 está uma representação simples de uma rede neuronal de retro-propagação.

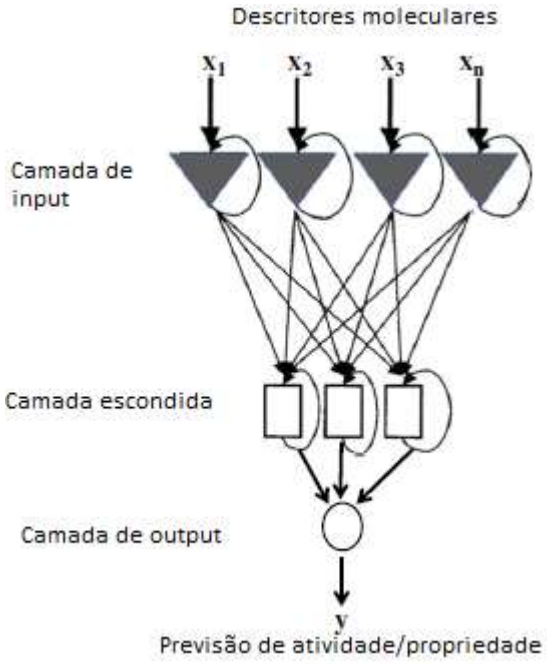

Figura 1.7 Representação de uma rede neuronal[30]

Em 1989 Rumelhart e a sua equipa desenvolveram um algoritmo que permitia o treino de redes neuronais de retro-propagação (BPANN).[43] As BPANN funcionam de forma a que com cada iteração, o modelo verifica os erros e aplica correções, alterando os pesos entre os neurónios, do fim para o início, de forma a minimizar o erro.

1.3.4 **Descritores moleculares**

Descritores moleculares são formas de representar moléculas em formatos que programas computacionais conseguem processar. Eles podem ser de 0D, 1D, 2D ou 3D em que D é dimensão. Na Tabela 1.1 encontram-se exemplos de descritores utilizados neste trabalho.

Dimensão	Descritor	Exemplos	
0D	Peso molecular, contagem de átomos, contagem de ligações	nC	Contagem do número de carbonos
		nAromBond	Contagem do número de ligações aromáticas
		nBondsD2	Número total de ligações duplas excluindo ligações aromáticas
		Sv	Soma de volumes van der Waals
		nHBAcc	Número de aceitadores de ligação de hidrogénio
1D	Contagem de fragmentos	C1SP3	Contagem de carbonos primários SP^3
		nAtomLC	Número de átomos na cadeia mais comprida
		MPC2	Total
		nRing	Número de aneis
2D	Descritores topológicos	ATS0m	Autocorrelação Broto-Moreau para massa
		BCUTw-1l	Descritor BCUT por peso atómico
		MDEC-11	Percursos entre carbonos primários
		WPATH	Número de caminhos Wiener

Tabela 1.1 Exemplos de descritores molecular utilizados[44]

Neste trabalho foram testados, mas não utilizados, descritores 3D por se verificar que não adicionavam informação útil para a construção dos modelos.

1.3.4.1 **Descritores 0D, 1D e 2D**

Descritores 1D e 2D retiram informação a partir da estrutura química a 2D. Por exemplo, no caso do 3-metilpentan-2-ol (fig. 1.7)

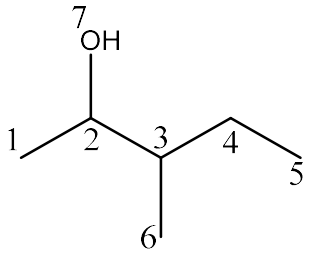

Figura 1.8 Representação do 3-metilpentan-2-ol

Um descritor 0D seria o número de carbonos na molécula: 6. Um descritor 1D podia ser o número de carbonos sp^3, que neste caso seria também 6.

Tomando o índice de Wiener como um exemplo de um descritor 2D, é possível obter a seguinte matriz de conectividade (tabela 1.2) para a molécula do 3-metilpentan-2-ol.

Tomando a seguinte matriz de conectividade (tabela 1.2) para a molécula do 3-metilpentan-2-ol, onde estão especificadas as distâncias entre átomos, o índice de Wiener pode ser calculado, como um exemplo de um descritor 2D.

	C1	C2	C3	O7	C4	C6	C5
C1	0	1	2	2	3	3	4
C2	1	0	1	1	2	2	3
C3	2	1	0	2	1	1	2
O7	2	1	2	0	3	2	1
C4	3	2	1	3	0	2	1
C6	3	2	1	2	2	0	3
C5	4	3	2	1	1	3	0

Tabela 1.2 Matriz de conectividade para o 3-metilpentan-2-ol

O índice de Wiener e definido pela equação:

$$W(G) = \frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1 \\ i\neq j}}^{N} d_{i,j} \quad (6)$$

onde *N* é o número de átomos na molécula, e $d_{i,j}$ representa a distância entre os átomos *i* e *j*. O índice de Wiener pode ser obtido somando as distâncias acima da diagonal principal da matriz de conectividade para uma molécula. Para o 3-metilpentan-2-ol, o W(G) = 42.

Um outro descritor utilizado são os vetores de autocorrelação 2D, dados pela equação:

$$a(d) = \sum_{i=1}^{N}\sum_{j=1}^{N} \delta(d_{i,j} - d)p_i p_j \quad (7)$$

Onde d representa a distância a considerar, $d_{i,j}$ a distância entre os átomos i e j, e p_x o valor da propriedade do átomo i ou j, respetivamente.

No caso mais simples, um vetor de autocorrelação usado é o número de par de átomos a várias distâncias, em que se considera que p = 1. Uma forma de calcular é construir uma tabela de conectividade e somar os pares de átomos a cada distância d que estão acima da diagonal principal. Tomando o 3-metilpentan-2-ol como exemplo, o a(3) = 5, visto que há 5 pares de átomos com uma distância de 3 acima da diagonal principal.

1.4 **Estado da arte**

Como já referido, a previsão de propriedades através de modelos QSPR está a ser uma técnica cada vez mais importante na química. No desenvolvimento de corantes para células solares já existem alguns trabalhos feitos.

Jie Xu et al. desenvolveram um modelo de previsão do máximo de absorção de corantes para DSSCs utilizando uma rede neuronal.[41]

Hongzhi Li et al. recorreram a um modelo de SVM em cascata para prever o PCE de corantes.[32] Para este modelo, os autores trabalharam com um conjunto de maioritariamente arilaminas. O modelo final obteve um R^2 de 0,75 e um RMSE de 0,78 na previsão da percentagem do PCE.

Supratik Kar et al. utilizaram regressão multilinear (MLR) para desenvolver um modelo de previsão do PCE[31] utilizando métodos computacionais chamados *density functional theory* (DFT) e *time dependent-DFT* (TD-DFT) para gerar os descritores. Separando os corantes por família, conseguiram valores de R^2 a variar entre 0,56 e 0,97 para o conjunto de teste.

2. Materiais e métodos computacionais

2.1 Conjunto de dados

Foi usada uma base de dados de 4426 corantes para células fotovoltaicas,[25] acedia em 17 de setembro 2018, existente na literatura. Começou-se por normalizar os *SMILES* de todas as moléculas utilizando o programa Standardizer da Chemaxon.[45,46] O processo de normalização consistiu na utilizando das opções indicadas na tabela 2.1.

Função	Objetivo ou Ação
Clean 2D	Calcula as posições de cada átomo na molécula
Remove Fragment	Remove os fragmentos, sendo escolhido o maior para ser mantido
Mesomerize	Transforma as moléculas para uma forma ressonante canónica
Aromatize	Aromatiza as moléculas, para todas ficarem com a mesma representação
Transform Aromatic N-Oxide	Transforma cada grupo funcional mencionado para garantir que cada grupo seja representado da mesma forma no SMILES
Transform Azide	
Transform Diazo	
Transform Diazonium	
Transform Iminium	
Transform Isocyanate	
Transform Nitrilium	
Transform Nitro	
Transform Nitrone Nitronate	
Transform Nitroso	
Transform Phosphonic	
Transform Phosphonium Ylide	
Transform Selenite	
Transform Silicate	
Transform Sulfine	
Transform Sulfon	
Transform Sulfonium Ylide	
Transform Sulfoxide	
Transform Sulfoxonium Ylide	
Transform Tertiary N-Oxide	
Add Explicit Hydrogens	Adiciona hidrogénios a estrutura, necessário para o PaDEL calcular os descritores

Tabela 2.1 Funções de normalização utilizadas

Seguidamente selecionaram-se somente os corantes que não são complexos de ruténio ou zinco, ficando com um conjunto de 3433 moléculas. Seguidamente foi feito uma seleção com base no eletrólito, focando-se no par redox I^-/I_3^- e selecionando somente as moléculas que continham a descrição completa da composição do eletrólito, passando assim a um conjunto com 2806 moléculas.

Do conjunto de 2806 moléculas, foi selecionado o subconjunto de 933 moléculas, todas com LiI+DMPII como eletrólito. Por incompatibilidade com o software utilizado, 51 moléculas que continham boro tiveram de ser removidas, chegando ao conjunto final de 882 moléculas. No final há 427 moléculas únicas, havendo 158 conjuntos de entre 2 e 17 moléculas repetidas.

Das 882 moléculas, 32 não têm os valores de máximo de absorção. Para o programa usado não pode haver valores em falta, portanto deu-se o valor de zero a esse parâmetro.

O conjunto final consiste em 882 sistemas experimentais definidos pela molécula do corante, comprimento de onda de máxima absorção, concentração de iodeto orgânico, concentração de iodeto inorgânico, concentração de t-butil piridina e outros aditivos, composição de solventes, área ativa da célula, concentração de co-adsorventes, espessura do filme de titânio, espessura da camada de *scattering*. A cada sistema estão associados valores experimentais de PCE, V_{OC}, J_{SC} e FF.

De cada conjunto, foi selecionado um conjunto de teste, que não foi usado para o treino dos modelos. O conjunto de teste foi composto ordenando de menor a maior o atributo a prever e selecionando aleatoriamente 10-15% das moléculas. Se foi selecionada uma molécula pertencente a um dos conjuntos de moléculas repetidas, esse conjunto passou todo o conjunto de teste.

2.2 **Descritores moleculares**

Da base de dados também foram tirados alguns dados para servirem de descritores do sistema: comprimento de onda de máxima absorção; concentração de iodeto orgânico; concentração de iodeto inorgânico; concentração de t-butil piridina e outros aditivos; composição de solventes; área ativa da célula; concentração de co adsorventes; espessura do filme de titânio; espessura da camada de *scattering*.

A estrutura molecular do corante foi representada pelos seguintes descritores moleculares.

Para cálculo do momento dipolar utilizou-se uma metodologia já existente,[35] e os calculados pelo *cxcalc* do programa *JChemSuite.*[45]

O cálculo de energias de orbitais moleculares foi feito seguindo metodologias já desenvolvidas[34], havendo 2 valores de energias HOMO (obtidas por métodos diferentes) e 3 valores de energias LUMO para cada molécula calculada.

Inicialmente foram calculados descritores com o programa CDK[47] e usando o programa WEKA[48] para os cálculos iniciais. Como os resultados não foram muito promissores, tentou-se outra metodologia. Usou-se o PaDEL[44] para calcular descritores. Foram calculados inicialmente descritores 0D, 1D e 2D, com a exceção do descritor FMF, totalizando 1443. Os descritores 3D foram calculados depois para comparação, sendo que as estruturas 3D foram otimizadas como descrito para o cálculo do momento dipolar.[35]

O descritor FMF mede a complexidade da molécula. Ele não foi calculado por se ter revelado demorado dado o grande conjunto de moléculas a estudar, mesmo quando se reduziu o grupo de moléculas a calcular em um quarto do número original.

2.3 **Descritores Correlacionados**

Para testar se os descritores correlacionados estavam a causar ruído, utilizou-se uma função de programas R.[49] Essa função calcula o coeficiente de correlação entre cada elemento, e depois se o valor absoluto da correlação for superior a um valor escolhido, retém o primeiro descritor e elimina o segundo. Essa seleção foi feita para um coeficiente de correlação de 0,8.

2.4 **Métodos de aprendizagem automática**

Foram construídos modelos de aprendizagem automática usando vários algoritmos diferentes, para prever os valores de PCE, V_{OC}, J_{SC} e FF dum sistema a partir da estrutura molecular do corante e dos parâmetros experimentais que definem o sistema.

2.4.1 **Random Forest**

Para contruir os modelos de regressão foi utilizado o programa R[50] e vários pacotes.[49,51] O R tem vantagens sobre o WEKA para o cálculo de *Random Forest* (RF), sendo que além de fornecer as previsões individuais em *out of bag* (OOB) para o conjunto de treino, também calcula a importância dos descritores.

O pacote utilizado para o R baseia-se nas florestas de árvores aleatórias desenvolvidas por Breiman.[33,51] Para construir os modelos, foi necessário especificar o m_{try} e o número de árvores a crescer. Utilizou-se sempre o mesmo número de árvores, 500, para cada modelo. O valor de m_{try} utilizado foi sempre um terço do número de descritores do conjunto a analisar.

Para construir cada modelo, foi criado um *script*, para facilmente alterar os valores dos parâmetros e o nome dos ficheiros de treino, teste e resultados.

2.4.2 **Rede neuronal**

Para a construção dos modelos de regressão utilizou-se o WEKA (versão 3.8.3).[48]

Para a construção da ANN, utilizou-se um protocolo já desenvolvido. Primeiro escolhe-se o método de *AdditiveRegression*[52] com 5 interações, seguido de *Bagging*[53] com 75 interações, e finalmente o *MLPRegressor* para a construção da rede neuronal com uma camada escondida.

O processo de *AdditiveRegression* com 5 iterações constrói 5 modelos, sendo que cada um corrige o erro do modelo anterior.

A função de *Bagging* com 75 iterações treina 75 modelos, separando os dados em 75 conjuntos diferentes, sendo o resultado a média dos resultados.

Para a construção dos modelos ANN os descritores foram normalizados utilizando a função de normalização de dados disponível no WEKA.

2.5 **Regressão**

Cada conjunto foi sempre separado em dois conjuntos, o conjunto de treino do modelo e o conjunto de teste do modelo. Existindo moléculas repetidas garantiu-se que essas moléculas estavam todas sempre presentes num só conjunto, ou o de treino ou teste.

Para avaliar os modelos, foi calculado um parâmetro, *Root-Mean-Square Error* (RMSE), que avalia o erro do *output* do modelo. O RMSE pode ser calculado segundo a seguinte equação:

$$RMSE = \sqrt{MSE} \tag{8}$$

O MSE pode ser calculado usando:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - x_i)^2 \tag{9}$$

onde *N* é o número total de elementos do conjunto, y_i o valor experimental e x_i o valor do modelo para o elemento *i*.

Para o conjunto de treino foram usados os valores de OOB e para o conjunto de teste foi usado o *output*. Também foi comparado o coeficiente de correlação entre o *output* e o valor experimental para cada molécula.

3. Resultados e discussão

Cada tabela apresenta os resultados da previsão de cada parâmetro, PCE, V_{OC}, J_{SC} e FF para cada conjunto de treino. Cada conjunto está representado da seguinte forma: classe de moléculas / descritores utilizados / RF conjunto de validação.

A classe de moléculas pode ser Todos (conjunto de todas as moléculas) (tabela 3.1) ou de uma só classe, por exemplo cumarina (para as moléculas de cumarina), tBP (dados que tenham só tBP como aditivo) ou tBP+*scattering* (dados que tenham tBP e camada de *scattering*) (tabelas 3.2).

Os descritores utilizados são os seguintes: Dm (1443 descritores moleculares do PaDEL), Dbd (propriedades experimentais retiradas da base de dados), Dom (estimativas de energias das orbitais moleculares), Dmd (estimativas do momento dipolar), 3D (descritores 3D).

No modelo Ts2 (tabela 3.2) foi testado substituir o comprimento de onda de máxima absorção (λ_{max}) pela distância desse máximo a 505 nm ($\Delta\lambda_{max}$).

O modelo excluindo correlações (Tabela 3.4) foram feitos removendo os descritores inter-correlacionados acima de 0,8.

A validação foi efetuada pelo erro em *out-of-bag* (OOB) durante o treino da RF, ou por previsão do conjunto de teste independente.

3.1 Modelos

3.1.1 Modelos utilizando o conjunto global

Inicialmente construíram-se modelos utilizando o conjunto total de moléculas. Em cada experiência foi-se acrescentando informação sobre os corantes/sistema de forma a avaliar a utilidade dessa informação para a previsão (Tabela 3.1).

Parâmetro	Nº de objetos	PCE		V_{OC}		J_{SC}		FF	
Conjunto		R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
T1 / Dm / RF OOB	772	0,6129	1,331	0,6372	65,608	0,5894	2,721	0,4624	0,0514
T1 / Dm / RF Teste	110	0,6053	1,337	0,6393	68,512	0,4597	3,127	0,4150	0,0490
T2 / Dm + Dbd / RF OOB	772	0,6478	1,260	0,7027	59,779	0,6212	2,618	0,6135	0,0419
T2 / Dm + Dbd / RF Teste	110	0,6597	1,261	0,7068	62,186	0,5140	2,986	0,5943	0,0436
T3 / Dm + Dbd + Dmd + Dom / RF OOB	772	0,6701	1,250	0,7058	60,304	0,5065	2,643	0,5880	0,0422

Parâmetro	Nº de objetos	PCE		V_{OC}		J_{SC}		FF	
Conjunto		R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
T3 / Dm + Dbd + Dmd + Dom / RF Teste	110	0,6542	1,244	0,6964	61,904	0,6140	3,005	0,6095	0,0437
T4 / Dm + 3D / RF OOB	772	0,5870	1,291	0,6194	64,929	0,5627	2,649	0,3190	0,0515
T4 / Dm + 3D / RF Teste	110	0,5443	1,283	0,6938	57,472	0,4785	2,863	0,5910	0,0481
T5 / Dm + Dbd + Dmd + Dom + 3D / RF OOB	772	0,6306	1,222	0,6773	59,831	0,5973	2,549	0,5308	0,0429
T5 / Dm + Dbd + Dmd + Dom + 3D / RF Teste	110	0,5211	1,311	0,7177	55,200	0,5228	2,755	0,7474	0,0410

Tabela 3.1 Modelos construídos recorrendo ao conjunto total de moléculas. Dm – Descritores moleculares, Dbd – propriedades retiradas da base de dados, Dom – orbitais moleculares, Dmd – momento dipolar, 3D – descritores 3D

Com base nos resultados concluiu-se que os descritores experimentais da base de dados, com especial relevo o máximo de absorção, adicionavam informação importante para a previsão (comparar resultados entre os modelos T1 e T2). Os descritores de momento dipolar e as energias das orbitais moleculares não revelaram adicionar informação útil na previsão (comparar modelos T2 e T3).

Os modelos (T4 e T5) serviram para testar se os descritores moleculares 3D adicionavam informação útil para a previsão - concluiu-se que aumentam muito o número de descritores sem trazerem aos resultados alterações significativas.

3.1.2 **Modelos utilizando subconjuntos**

Seguidamente fizeram-se várias experiências restringindo o conjunto de dados a subconjuntos com alguma característica comum, plausivelmente relevante para o comportamento do sistema. O modelo da cumarina (C) foi feito na hipótese que a separação das várias famílias de corantes podia dar melhores resultados. Na sequência do modelo da cumarina foi feita outra separação, neste caso sendo entre um conjunto de três fatores: se o teste do corante incluía t-butil piridina no eletrólito na ausência (modelos Tp1 e Tp2) ou presença (modelos Ts1, Ts2 e Ts3) da camada de *scattering*. No modelo Ts2 foi feito um teste, o máximo de absorção foi alterado pela distância do máximo de absorção a 505 nm.

Parâmetro	Nº de objetos	PCE		V_{OC}		J_{SC}		FF	
Conjunto		R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
C / Dm + Dbd + Dmd + Dom / RF OOB	**61**	0,7680	0,953	0,7719	45,502	0,5019	2,237	0,8368	0,0389
C / Dm + Dbd + Dmd + Dom / RF Teste	**12**	0,6815	1,613	0,5989	46,536	0,8272	2,690	0,4471	0,0631
Tp1 / Dm + Dmd + Dom / RF OOB	**253**	0,5600	1,158	0,5110	66,145	0,4274	2,597	0,2930	0,0521
Tp1 / Dm + Dmd + Dom / RF Teste	**33**	0,4198	1,911	0,5823	68,085	0,3298	4,022	0,5282	0,0475
Tp2 / Dm + Dmd + Dom + λ_{max} / RF OOB	**203**	0,4210	1,344	0,4463	72,672	0,2720	3,025	0,3328	0,0534
Tp2 / Dm + Dmd + Dom + λ_{max} / RF Teste	**25**	0,6684	1,471	0,7275	64,467	0,5410	3,278	0,1399	0,0626
Tp3 / Dm + Dmd + Dom + Dbd / RF OOB	**253**	0,6086	1,097	0,5339	64,674	0,4782	2,484	0,3243	0,0509
Tp3 / Dm + Dmd + Dom + Dbd / RF Teste	**33**	0,3987	1,920	0,5584	69,091	0,3433	3,966	0,5477	0,0715
Ts1 / Dm + Dmd + Dom + λ_{max} / RF OOB	**182**	0,6210	1,141	0,6035	41,527	0,6235	2,258	0,4998	0,0316
Ts1 / Dm + Dmd + Dom + λ_{max} / RF Teste	**21**	0,6170	1,345	0,5520	40,297	0,5155	2,659	0,4059	0,0394
Ts2 / Dm + Dmd + Dom + $\Delta\lambda_{max}$ RF OOB	**182**	0,6223	1,343	0,6016	49,139	0,5970	2,714	0,5035	0,0371
Ts2 / Dm + Dmd + Dom + $\Delta\lambda_{max}$ / RF Teste	**21**	0,5702	1,518	0,5367	44,836	0,5256	2,856	0,3389	0,0455
Ts3 / Dm + Dmd + Dom / RF OOB	**182**	0,6154	1,349	0,6076	48,702	0,5882	2,728	0,4900	0,0375
Ts3 / Dm + Dmd + Dom / RF Teste	**21**	0,5515	1,565	0,5203	46,097	0,4828	2,988	0,3991	0,0435
Ts4 / Dm + Dbd + Dmd + Dom / RF OOB	**182**	0,6160	1,832	0,6112	48,619	0,6274	2,639	0,5289	0,0361
Ts4 / Dm + Dbd + Dmd + Dom / RF Teste	**21**	0,6009	1,479	0,5396	44,718	0,5451	2,816	0,6024	0,0374

Tabela 3.2 Modelos construídos recorrendo a subconjuntos de moléculas. Dm – Descritores moleculares, Dbd – propriedades retiradas da base de dados, Dom – orbitais moleculares, Dmd – momento dipolar, λ_{max} – comprimento de onda máximo, $\Delta\lambda_{max}$ – distancia entre 505 nm e λ_{max}

3.1.2.1 **Modelo das cumarinas**

Para comparar o modelo das cumarinas, foram retirados os resultados do modelo T3 só para as moléculas de cumarina (tabela 3.3).

Parâmetro	PCE		VOC		JSC		FF	
Conjunto	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
T3 / Dm + Dbd + Dmd + Dom / RF OOB	0,7730	0,927	0,7205	46,552	0,5641	2,050	0,8470	0,0374
C / Dm + Dbd + Dmd + Dom / RF OOB	0,7680	0,953	0,7719	45,502	0,5019	2,237	0,8368	0,0389
T3 / Dm + Dbd + Dmd + Dom / RF Teste	0,6542	1,244	0,6964	61,904	0,6140	3,005	0,6095	0,0437
C / Dm + Dbd + Dmd + Dom / RF Teste	0,6815	1,613	0,5989	46,536	0,8272	2,690	0,4471	0,0631

Tabela 3.3 Comparação de resultados para moléculas de cumarina Dm – Descritores moleculares, Dbd – propriedades retiradas da base de dados, Dom – orbitais moleculares, Dmd – momento dipolar

Verifica-se que usando apenas sistemas com corantes de cumarina o poder de previsão do piora. Existe um aumento médio do erro de 3,4% na previsão OOB e de 9,69% na previsão do conjunto de teste. Conclui-se assim que separar as moléculas por família não é um passo importante na construção do modelo.

3.1.2.2 **Modelo t-butil piridina (Tp)**

Como fui concluído que separar o conjunto em conjuntos de famílias de moléculas, foi testado outra separação, sendo uma dessas a criação de um subconjunto de moléculas em que os dados experimentais foram obtidos na presença de t-butil piridina.

O primeiro modelo (Tp1) foi construído utilizando os descritores moleculares, descritores da previsão das orbitais moleculares e descritores da previsão do momento dipolar. Os descritores da base de dados foram omitidos, pois neste caso seriam iguais para todos os sistemas.

O modelo Tp2 foi construído utilizando o mesmo subconjunto, mas com a inclusão do máximo de absorção do corante, onde foram excluídas as moléculas onde esse dado não era conhecido. Este teste permitiu concluir que o máximo de absorção pode ser um descritor importante, pois enquanto que o erro médio da previsão OOB aumenta 11,2%, o erro médio da previsão do conjunto de teste diminui em 3,7% (ver tabela 3.2). No entanto, o máximo de absorção nem sempre é conhecido, visto que é uma propriedade obtida experimentalmente (e se a molécula já existir) ou prevista por algum modelo.

O modelo Tp3 foi construído com a adição dos descritores de base de dados. Quando comparando com o modelo Tp1, verifica-se um decréscimo do erro médio na previsão do OOB de 3,5% e um aumento médio na previsão do conjunto de teste de 12,8%.

3.1.2.3 **Modelo t-butil piridina e scattering (Ts)**

Para testar com outro subconjunto, foram separadas as moléculas onde foi utilizado t-butil piridina e existe uma camada de *scattering* na célula.

O primeiro modelo (Ts1) foi construído da mesma forma que o modelo Tp2, descritores moleculares, descritores da previsão das orbitais moleculares, descritores da previsão do momento dipolar e o máximo de absorção do corante.

O modelo Ts2 foi testado a distância do máximo de absorção a 505nm. Neste caso, comparando com o modelo Ts1, há um aumento médio de 18,4% no erro do OOB e um aumento médio de 11,8% do erro do conjunto de teste. Este resultado mostra que não é a distância ao máximo do espectro solar em si que é importante neste descritor. Um descritor melhor poderia ser a percentagem de sobreposição do espectro de absorção do corante com o espectro de emissão solar.

O modelo Ts3 foi construído pelas razões mencionadas acima, pois não inclui o máximo de absorção. Neste caso, a ausência desse descritor aumentou o erro médio OOB em 18,7% e o erro médio do conjunto de teste em 13,4%.

No modelo Ts4 foi testado se os descritores da base de dados (Dbd) têm importância na previsão. Comparando com o modelo Ts1, o erro médio na previsão aumento 27,2% no conjunto OOB e 5,4% conjunto de teste.

3.1.3 **Modelo sem correlações (Cor)**

Parâmetro		**PCE**		**V_{OC}**		**J_{SC}**		**FF**	
Conjunto	**Nº de objetos**	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
Cor / Dm + (Dbd – Cor) / RF OOB	**772**	0,6712	1,223	0,7071	59,302	0,6425	2,554	0,6244	0,0414
Cor / (Dm + Dbd) – Cor / RF Teste	**110**	0,6596	1,267	0,7130	61,740	0,5277	2,950	0,5871	0,0433

Tabela 3.4 Resultados do modelo RF para todas as moléculas utilizando os descritores do PaDEL (Dm) e descritores da base de dados (Dbd) removendo os descritores corelacionados (– Cor)

O modelo sem correlações foi construído de modo a retirar os descritores correlacionados entre si. Este teste serviu para ver se é necessário calcular todos os descritores e se o ruído ou informação redundante dos vários descritores estava a afetar a construção dos modelos. Comparando os resultados com os obtidos no modelo T2 (Tabela 3.1) verificou-se que para o modelo RF a presença de descritores altamente

corelacionados não tem impacto nos resultados. Para cada parâmetro calculado, as diferenças no RMSE são muito baixas, podendo considerar-se que não há variação significativa.

3.1.4 **Modelo NN usando 20 descritores selecionados via RF (NN)**

Para além dos modelos treinados com RF indicados até aqui, foi também explorado um algoritmo de redes neuronais artificiais. O modelo utilizado contém uma camada de entrada com 20 nodos, uma camada escondida com 6 nodos e uma camada de saída com um nodo. A função de ativação dos nodos da camada escondida utilizada foi a *ApproximateSigmoid*, sendo que no nodo de saída, no caso de regressão, a função utilizada foi a função identidade.

Parâmetro		**PCE**		**V_{OC}**		**J_{SC}**		**FF**	
Conjunto	**Nº de objetos**	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
Todos / NN com 20 descritores / NN Treino	**772**	0,6237	1,299	0,6640	69,146	0,5969	2,692	0,6071	0,0423
Todos / NN com 20 descritores / NN Teste	**110**	0,6857	1,210	0,6232	63,755	0,4893	3,062	0,5309	0,0447

Tabela 3.5 Resultados dos modelos de ANN para todas as moléculas usando os 20 descritores mais importantes recorrendo aos resultados do modelo T3

Cada modelo de previsão dos parâmetros (PCE, V_{OC}, J_{SC} e FF) foi construído usando os 20 descritores mais importantes selecionados na construção do modelo T3 de cada parâmetro a prever.

Comparando com a tabela 3.6 é possível ver que utilizando só 20 descritores foi possível chegar a resultados semelhantes aos da melhor RF.

3.2 **Resultados globais**

Olhando para os resultados dos vários modelos, é possível ver que os modelos conseguem prever, em média, os valores do conjunto de teste com alguma confiança. O melhor modelo conseguido foi o modelo Ts1. Este modelo pode ser utilizado para a previsão de corantes na presença de t-butil piridina e com camada de *scattering* na célula.

De modo geral os modelos conseguem distinguir os corantes com parâmetros baixos e altos. No entanto os modelos não conseguem fazer distinção fina entre os corantes, levando a uma dispersão dos resultados (fig. 3.1).

Parâmetro	PCE		V_{OC}		J_{SC}		FF	
Conjunto	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
Ts1 / Dm + Dmd + Dom + λ_{max} / RF OOB	0,6210	1,141	0,6035	41,527	0,6235	2,258	0,4998	0,0316
Ts1 / Dm + Dmd + Dom + λ_{max} / RF Teste	0,6170	1,345	0,5520	40,297	0,5155	2,659	0,4059	0,0394

Tabela 3.6 Modelo Ts1

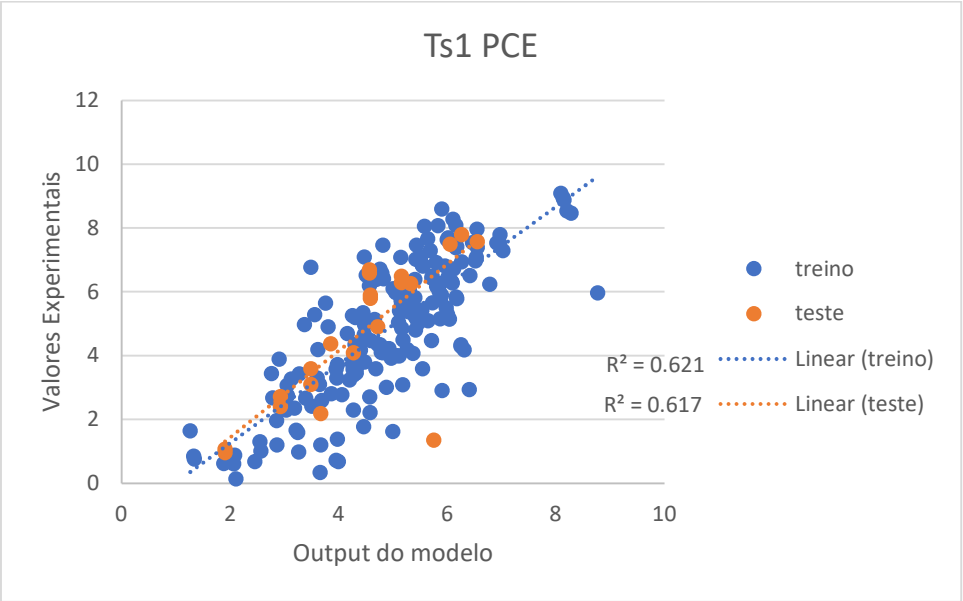

Figura 3.1 Resultados para o PCE do modelo Ts1

3.3 Analise da importância dos descritores para o V_{OC}

Os dez descritores mais importantes na previsão do VOC para o modelo Ts1 são os seguintes:

Descritor	Propriedade
nAtomLAC	Comprimento da cadeia alifática mais comprida
AATSC3s	Autocorrelação Broto-Moreau centrada média ponderada para o estado intrínseco
ATSC4i	Autocorrelação Broto-Moreau centrada ponderada para o primeiro potencial de ionização
minHCsats	O valor mínimo do estado electro topológico mínimo de hidrogénios ligados a carbonos saturados
GATS3c	Autocorrelação Geary ponderada por cargas
GATS2p	Autocorrelação Geary ponderada por polarizabilidades
GATS2m	Autocorrelação Geary ponderada por massa
MATS3s	Autocorrelação Moran ponderada para o estado intrínseco
AATSC0i	Autocorrelação Broto-Moreau centrada média ponderada para o primeiro potencial de ionização
WTPT-5	Soma de comprimentos de caminhos começando em azoto

Tabela 3.7 Os dez descritores mais importantes para o V_{OC} do modelo Ts1

Analisando os resultados da tabela 3.7 existem descritores ligados a propriedades eletrónicas. Para além disso, a importância de dois descritores merece uma nota.

O primeiro é o NAtomLAC, que conta o caminho mais comprido feito de carbonos alifáticos. Estes grupos são usados para impedir agrupamentos muito compactos de corante à superfície do semicondutor.

O segundo é o WTPT-5. Este descritor é a soma dos comprimentos de caminhos começando em azoto. Um grupo de ancoragem comum é o acido ciano acrílico, em que o azoto é o doador do eletrão para o semicondutor.

3.4 Analise da importância dos descritores para o J_{SC}

Os dez descritores mais importantes para a previsão do J_{SC} no modelo Ts1 são os seguintes:

Descritor	Propriedade
Molecule spectrum absorption maxima	Máximo de absorção do corante
ATSC4s	Autocorrelação Broto-Moreau centrada ponderada para o estado intrínseco
MATS6c	Autocorrelação Moran ponderada por cargas
GATS6c	Autocorrelação Geary ponderada por cargas
StN	Soma dos estados electro topológicos de azotos com tripla ligação
maxtN	Estado electro topológico máximo de azotos com ligação tripla
ATSC6c	Autocorrelação Broto-Moreau centrada ponderada por cargas
MAXDP	Máxima diferença positiva de estados intrínsecos da molécula (relacionado com a eletrofilicidade)
MATS1c	Autocorrelação Moran ponderada por cargas
VR2_D	Índice baseado no vetor próprio do tipo Randic normalizado da matriz de distância topológica

Tabela 3.8 Os dez descritores mais importantes para o JSC do modelo Ts1

Na tabela 3.8 estão escritores ligados a propriedades eletrónicas. No entanto há um foco no máximo de absorção do corante, que se revelou o mais importante na previsão do J_{SC}. Existem também dois descritores ligados a azotos com ligação tripla, mostrando a relevância da presença do grupo de ancoragem - acido ciano acrílico.

3.5 Analise da importância dos descritores para o FF

Os dez descritores mais importantes para a previsão do FF no modelo Ts1 são os seguintes:

Descritor	Propriedade
WTPT-5	Soma de comprimentos de caminhos começando em azoto
AATS7v	Autocorrelação Broto-Moreau média ponderada por volumes de van der Waals
AATS7p	Autocorrelação Broto-Moreau média ponderada por polarizabilidades
GATS8i	Autocorrelação Geary ponderada pelo primeiro estado de ionização
AATSC8v	Autocorrelação Broto-Moreau centrada média ponderada por volumes de van der Waals
MATS8v	Autocorrelação Moran ponderada por volumes de van der Waals
SpMax1_Bhe	Vetor próprio máximo, em valor absoluto, da matriz de Burden modificada ponderado pelas eletronegatividades de Sanderson relativas
MATS8p	Autocorrelação Moran ponderada por polarizabilidades
GATS8m	Autocorrelação Geary ponderada pela massa
SpMax2_Bhm	Vetor próprio máximo, em valor absoluto, da matriz de Burden modificada ponderado pela massa relativa

Tabela 3.9 Os dez descritores mais importantes para o FF do modelo Ts1

Olhando para a tabela 3.9 nota-se que há uma mistura de propriedades físicas e eletrónicas nos descritores mais importantes. A previsão do FF foi das mais inconstantes, e os descritores mais importantes variaram muito entre modelos, apontando para a possibilidade de haver muito ruído envolvido.

3.6 **Analise da importância dos descritores para o PCE**

Os dez descritores mais importantes para a previsão do FF no modelo Ts1 são os seguintes:

Descritor	Propriedade
ATSC4s	Autocorrelação Broto-Moreau centrada ponderada para o estado intrínseco
ETA_dBeta	Uma medida do conteúdo insaturado relativo
MATS1c	Autocorrelação Moran ponderada por cargas
MATS6c	Autocorrelação Moran ponderada por cargas
maxtN	Estado electro topológico máximo de azotos com ligação tripla
MATS1s	Autocorrelação Moran ponderada para o estado intrínseco
ETA_BetaP_ns_d	Uma medida de eletrões solitários a entrar em ressonância relativo ao tamanho molecular
HybRatio	Ratio de carbonos sp3 e carbonos sp2
ATSC6c	Autocorrelação Broto-Moreau centrada ponderada por cargas
Molecule spectrum absorption maxima	Máximo de absorção do corante

Tabela 3.10 Os dez descritores mais importantes para o PCE do modelo Ts1

Olhando para a tabela 3.10, existe um foco em propriedades eletrónicas na previsão da eficiência. O máximo de absorção aparece como um descritor importante mais uma vez, tal como um descritor ligado a azotos com ligação tripla.

O PCE é um produto dos outros fatores (equação 1), e por isso sofre do mesmo problema de ruído que a previsão do FF tem. Quando se usa o modelo de descritores moleculares para prever, existe algum padrão, mas como na previsão do FF, existe muita variação entre modelos.

3.7 Comparação com modelos existentes na literatura

No artigo por Li et al.[32] foi utilizado o RMSE como avaliação dos modelos de arilaminas por SVM, podendo assim comparar-se os resultados obtidos para o PCE com o modelo Ts1 (tabela 3.2).

	Modelo das cumarinas		**Modelo do artigo[32]**	
	RMSE	R^2	RMSE	R^2
PCE	1,345	0,6170	0,78	0,75

Tabela 3.11 Comparação de resultados com a literatura

Comparando os resultados dos conjuntos de treino, verifica-se que os resultados deste trabalho são inferiores. O modelo aqui apresentado baseia-se simplesmente em descritores moleculares empíricos, rapidamente calculáveis, enquanto que Li et al. utilizaram cálculos por *density functional theory* (DFT) para obter os descritores. Esses cálculos podem dar descritores que descrevem melhor o funcionamento do corante, dando assim um modelo melhor, mas com a desvantagem de necessitar cálculos demorosos.

4. **Conclusões**

O objetivo deste trabalho foi construir modelos de previsão que dessem indicações sobre os parâmetros ideais a serem considerados no desenvolvimento de corantes para células fotovoltaicas. A utilização do modelo RF foi favorável, pois com os dados obtidos, foi possível calcular os descritores para as 882 moléculas e utilizá-los sem pré-seleção e sem se perder poder de previsão.

Conseguiu-se construir vários modelos com capacidade de dar uma ideia de como o corante se comporta, conseguindo diferenciar entre corantes com os parâmetros altos e baixos. Os valores previstos não são ideais, mas os modelos podem servir para fazer uma seleção entre corantes promissores e corantes não promissores, assim poupando tempo e recursos no laboratório.

Infelizmente não foi possível chegar a uma regra empírica sobre a construção de corantes. No entanto, a presença de vários vetores de autocorrelação pode ser indicador que o modelo está a diferenciar com base na presença ou não de alguns grupos funcionais, visto que esses grupos vão ter valores similares entre moléculas diferentes.

Em relação ao V_{OC}, um descritor molecular que aparece com alguma importância é o comprimento da cadeia alifática mais comprida na molécula. Este resultado vem em linha com resultados experimentais já obtidos, que indicam que uma cadeia alifática ou aumento da área apolar dos corantes aumenta o V_{OC}[15,18].

Olhando para os resultados da importância em relação ao J_{SC} é possível reparar que em geral, o máximo de absorção, momento dipolar e as energias das HOMO/LUMO têm algum impacto.

O máximo de absorção é importante por razões probabilísticas. Olhando para a equação 2, no cálculo do J_{SC} entra o termo que representa o fluxo de fotões ao comprimento de onda λ, $I_S(\lambda)$, do espetro de emissão do sol. O J_{SC} é calculado como a sobreposição entre o espetro de conversão para cada comprimento de onda do corante e o espetro solar. Sendo assim, existe uma maior probabilidade da transição HOMO/LUMO num corante em que essa sobreposição é maior, assim dando um valor de corrente maior.

O momento dipolar afeta a eficiência de injeção. Moléculas com um momento dipolar perpendicular ao semicondutor têm uma eficiência maior.

No que toca as energias das orbitais moleculares, em primeiro lugar é necessário que a LUMO do corante esteja acima da energia da banda condutora do semicondutor, e que a HOMO esteja abaixo da banda do par redox. No entanto, quanto maior a diferença entre a LUMO e a banda condutora, mais rápida será a transferência do eletrão. O mesmo princípio aplica-se a diferença entre à HOMO e o par redox.

Em relação ao *Fill Factor*, existe algum padrão entre os modelos. No entanto existem problemas em prever o valor do FF. Essa dificuldade existe, pois, além de fatores moleculares, o FF é muito afetado pelos materiais e a construção da célula. Devido a

isso, existe muito ruído, sendo de difícil interpretação os resultados da importância dos descritores na previsão do FF.

Para a previsão do *Fill Factor*, existem problemas relacionados com a construção da célula solar que podem influenciar esse parâmetro. Assim sendo propõe-se que a utilização de um corante padrão poderia ajudar. Usando um padrão seria possível utilizar o FF normalizado (FF_N), podendo ser calculado da seguinte forma:

$$FF_N = FF \times \frac{FF_M}{FF_P} \quad (10)$$

em que o FF_M é o FF do corante padrão medido nas mesmas condições do corante a medir e o FF_P é o FF do corante padrão.

A utilização da base de dados elaborada por Venkatraman et al.[25] foi muito útil para este trabalho, no entanto foram observados alguns problemas. Primeiramente alguns dos códigos SMILES das moléculas estavam incorretos e as estruturas tiveram de ser verificadas no artigo fonte. Um outro problema refere-se à inserção de dados. Em alguns casos houve parâmetros trocados ou então inseridos de forma diferente do restante. Por exemplo, na maioria dos casos o parâmetro do *fill factor* foi dado como um valor que varia entre 0 e 1, mas em alguns casos foi inserido como uma %, variando assim entre 0 e 100.

Um outro problema mais geral foi a falta de dados do *dye loading*. Esse parâmetro pode ser importante para a estimativa dos parâmetros fotovoltaicos da célula solar, pois mede a capacidade de adsorção do corante à superfície do semicondutor. Só 189 das 882 entradas tinham esse dado na base de dados.

5. Referencias

1. B.D. Solomon, K. Krishna. **The coming sustainable energy transition: History, strategies, and outlook**. *Energy Policy*. **2011**;*39*(11):7422-7431. doi:10.1016/j.enpol.2011.09.009

2. N. Kannan, D. Vakeesan. **Solar energy for future world: - A review**. *Renew Sustain Energy Rev*. **2016**;*62*:1092-1105. doi:10.1016/j.rser.2016.05.022

3. E. Burt, P. Orris, S. Buchanan. **Scientific Evidence of Health Effects from Coal Use in Energy Generation**. **2013**;(April):1-18.

4. P. Schou. **Polluting non-renewable resources and growth**. *Environ Resour Econ*. **2000**;*16*(2):211-227. doi:10.1023/A:1008359225189

5. N.L. Panwar, S.C. Kaushik, S. Kothari. **Role of renewable energy sources in environmental protection: A review**. *Renew Sustain Energy Rev*. **2011**;*15*(3):1513-1524. doi:10.1016/j.rser.2010.11.037

6. E. Kabir, P. Kumar, S. Kumar, et al. **Solar energy: Potential and future prospects**. *Renew Sustain Energy Rev*. **2018**;*82*(September 2017):894-900. doi:10.1016/j.rser.2017.09.094

7. S. Allen. **Carbon footprint of electricity gerenation**. *Parliam Off Sci Technol*. **2011**;*383*(383):1-4. https://www.parliament.uk/documents/post/postpn_383-carbon-footprint-electricity-generation.pdf.

8. EIA. **Net Generation by State by Type of Producer by Energy Source**. Annual Data for 2017. http://www.eia.gov/electricity/data/state/. Published 2019. Accessed April 4, 2019.

9. A. Lopez, B. Roberts, D. Heimiller, et al. *U.S. Renewable Energy Technical Potentials: A GIS-Based Analysis*.; 2012.

10. A. Carella, F. Borbone, R. Centore. **Research Progress on Photosensitizers for DSSC**. **2018**;*6*(October):1-24. doi:https://dx.doi.org/10.3389%2Ffchem.2018.00481

11. W. Shockley, H.J. Queisser. **Detailed Balance Limit of Efficiency of p-n Junction Solar Cells**. *J Appl Phys*. **1961**;*32*(3):510-519. doi:10.1063/1.1736034

12. M.A. Green, Y. Hishikawa, E.D. Dunlop, et al. **Solar cell efficiency tables (version 51)**. *Prog Photovoltaics Res Appl*. **2018**;*26*(1):3-12. doi:10.1002/pip.2978

13. **Best Research-Cell Efficiency Chart**. https://www.nrel.gov/pv/cell-efficiency.html. Accessed May 3, 2019.

14. B. O'Regan, M. Grätzel. **A low-cost, high-efficiency solar cell based on dye-sensitized colloidal TiO2 films**. *Nature*. **1991**;*353*(6346):737-740. doi:10.1038/353737a0

15. B.G. Kim, K. Chung, J. Kim. **Molecular design principle of all-organic dyes for dye-sensitized solar cells**. *Chem - A Eur J*. **2013**;*19*(17):5220-5230. doi:10.1002/chem.201204343

16. Y. Ooyama, Y. Harima. **Photophysical and electrochemical properties, and molecular structures of organic dyes for dye-sensitized solar cells**. *ChemPhysChem*. **2012**;*13*(18):4032-4080. doi:10.1002/cphc.201200218

17. L. Zhang, J.M. Cole. **Anchoring groups for dye-sensitized solar cells**. *ACS Appl Mater Interfaces*. **2015**;*7*(6):3427-3455. doi:10.1021/am507334m

18. Z. Ning, Y. Fu, H. Tian. **Improvement of dye-sensitized solar cells: What we know and what we need to know**. *Energy Environ Sci*. **2010**;*3*(9):1170-1181. doi:10.1039/c003841e

19. S.G. Hashmi, M. Özkan, J. Halme, et al. **Dye-sensitized solar cells with inkjet-printed dyes**. *Energy Environ Sci*. **2016**;*9*(7):2453-2462. doi:10.1039/C6EE00826G

20. L. El Chaar, L.A. Lamont, N. El Zein. **Review of photovoltaic technologies**. *Renew Sustain Energy Rev*. **2011**;*15*(5):2165-2175. doi:10.1016/j.rser.2011.01.004

21. P. Karthika, S. Ganesan, M. Arthanareeswari. **Low-cost synthesized organic compounds in solvent free quasi-solid state polyethyleneimine, polyethylene glycol based polymer electrolyte for dye-sensitized solar cells with high photovoltaic conversion efficiencies**. *Sol Energy*. **2018**;*160*(November 2017):225-250. doi:10.1016/j.solener.2017.11.076

22. J. Zhao, X. Shen, F. Yan, et al. **Solvent-free ionic liquid/poly(ionic liquid) electrolytes for quasi-solid-state dye-sensitized solar cells**. *J Mater Chem*. **2011**;*21*(20):7326-7330. doi:10.1039/c1jm10346f

23. Y. Li, X.-L. He, C.-X. Lian, et al. **In Situ Formation of Continuous Charge Transfer Pathways for Highly Efficient, Solvent-Free, Polymer Electrolyte-Based Dye-Sensitized Solar Cells**. *ACS Sustain Chem Eng*. **2016**;*4*(7):4013-4020. doi:10.1021/acssuschemeng.6b00908

24. J. Gong, K. Sumathy, Q. Qiao, et al. **Review on dye-sensitized solar cells (DSSCs): Advanced techniques and research trends**. *Renew Sustain Energy Rev*. **2017**;*68*(July 2016):234-246. doi:10.1016/j.rser.2016.09.097

25. V. Venkatraman, R. Raju, S.P. Oikonomopoulos, et al. **The dye-sensitized solar cell database**. *J Cheminform*. **2018**;*10*(1):1-9. doi:10.1186/s13321-018-0272-0

26. **Fill Factor**. https://www.pveducation.org/pvcdrom/solar-cell-operation/fill-factor. Accessed May 27, 2019.

27. A.F.A. Cros. **Action de L'alcool amylique sur L'organisme**. **1863**.

28. C. Hansch, P.P. Maloney, T. Fujita, et al. **Correlation of Biological Activity of Phenoxyacetic Acids with Hammett Substituent Constants and Partition Coefficients**. *Nature*. **1962**;*194*(4824):178-180. doi:10.1038/194178b0

29. C. Hansch, R.M. Muir, T. Fujita, et al. **The Correlation of Biological Activity of Plant Growth Regulators and Chloromycetin Derivatives with Hammett Constants and Partition Coefficients**. *J Am Chem Soc*. **1963**;*85*(18):2817-2824. doi:10.1021/ja00901a033

30. S. Yousefinejad, B. Hemmateenejad. **Chemometrics tools in QSAR/QSPR studies: A historical perspective**. *Chemom Intell Lab Syst*. **2015**;*149*:177-204. doi:10.1016/j.chemolab.2015.06.016

31. S. Kar, J.K. Roy, J. Leszczynski. **In silico designing of power conversion efficient organic lead dyes for solar cells using todays innovative approaches to assure renewable energy for future**. *npj Comput Mater*. **2017**;*3*(1):22. doi:10.1038/s41524-017-0025-z

32. H. Li, Z. Zhong, L. Li, et al. **A cascaded QSAR model for efficient prediction of overall power conversion efficiency of all-organic dye-sensitized solar cells**. *J Comput Chem*. **2015**;*36*(14):1036-1046. doi:10.1002/jcc.23886

33. L. Breiman. **Random Forests**. *Mach Learn*. **2001**;*45*(1):5-32. doi:10.1023/A:1010933404324

34. F. Pereira, K. Xiao, D.A.R.S. Latino, et al. **Machine Learning Methods to Predict Density Functional Theory B3LYP Energies of HOMO and LUMO Orbitals**. *J Chem Inf Model*. **2017**;*57*(1):11-21. doi:10.1021/acs.jcim.6b00340

35. F. Pereira, J. Aires-de-Sousa. **Machine learning for the prediction of molecular dipole moments obtained by density functional theory**. *J Cheminform*. **2018**;*10*(1). doi:10.1186/s13321-018-0296-5

36. V. Svetnik, A. Liaw, C. Tong, et al. **Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling**. *J Chem Inf Comput Sci*. **2003**;*43*(6):1947-1958. doi:10.1021/ci034160g

37. A. Liaw, M. Wiener. **Classification and Regression by randomForest**. *R News*. **2002**;*2*(3):18-22. https://cran.r-project.org/doc/Rnews/.

38. A. Lavecchia. **Machine-learning approaches in drug discovery: Methods and applications**. *Drug Discov Today*. **2015**;*20*(3):318-331. doi:10.1016/j.drudis.2014.10.012

39. A. Mellit, S.A. Kalogirou. **Artificial intelligence techniques for photovoltaic applications: A review**. *Prog Energy Combust Sci*. **2008**;*34*(5):574-632. doi:10.1016/j.pecs.2008.01.001

40. A. Mellit, S.A. Kalogirou, L. Hontoria, et al. **Artificial intelligence techniques for sizing photovoltaic systems: A review**. *Renew Sustain Energy Rev*. **2009**;*13*(2):406-419. doi:10.1016/j.rser.2008.01.006

41. J. Xu, H. Zhang, L. Wang, et al. **Artificial neural network-based QSPR study on absorption maxima of organic dyes for dye-sensitised solar cells**. *Mol Simul*. **2011**;*37*(1):1-10. doi:10.1080/08927022.2010.506513

42. W.S. McCulloch, W. Pitts. **A logical calculus of the ideas immanent in nervous activity**. *Bull Math Biol*. **1990**;*52*(1-2):99-115. doi:10.1007/BF02459570

43. D.E. Rumelhart, G.E. Hinton, R.J. Williams. **Learning representations by back-propagating errors**. *Nature*. **1986**;*323*(6088):533-536. doi:10.1038/323533a0

44. C.W. Yap. **PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints**. *J Comput Chem*. **2011**;*32*(7):1466-1474. doi:10.1002/jcc.21707

45. Chemaxon. **JChemSuite**. **2017**. http://www.chemaxon.com.

46. **Funções de normalização do Standardizer**. https://docs.chemaxon.com/display/docs/Standardizer+Actions. Accessed May 3, 2019.

47. E.L. Willighagen, J.W. Mayfield, J. Alvarsson, et al. **The Chemistry Development Kit (CDK) v2.0: atom typing, depiction, molecular formulas, and substructure searching**. *J Cheminform*. **2017**;*9*(1):1-19. doi:10.1186/s13321-017-0220-4

48. E. Frank, M.A. Hall, I.H. Witten. *The WEKA Workbench Data Mining: Practical Machine Learning Tools and Techniques*. 4th ed. Morgan Kaufmann; 2016. doi:10.1016/B978-0-12-804291-5.00024-6

49. M. Khun. **caret package for R**. **2018**. https://cran.r-project.org/package=caret.

50. R Core Team. **R: A language and environment for statistical computing. 2015**. https://www.r-project.org/.

51. A. Liaw, M. Wiener. **randomForest package for R**. **2018**. https://cran.r-project.org/package=randomForest.

52. J.H. Friedman. **Stochastic gradient boosting**. *Comput Stat Data Anal*. **2002**;*38*(4):367-378. doi:10.1016/S0167-9473(01)00065-2

53. L. Breiman. **Bagging predictors**. *Mach Learn*. **1996**;*24*(2):123-140. doi:10.1007/BF00058655

6. Anexos

6.1 Previsões OOB para o conjunto de treino do modelo Ts1

Molecule ID	Smiles	V_{OC}	Previsão V_{OC}	J_{SC}	Previsão J_{SC}	FF	Previsão FF	PCE	Previsão PCE
45	CC1(C)CCN2CCC(C)(C)c3c2c1cc1cc(-c2ccc(s2)-c2ccc(\C=C(\C#N)C(O)=O)s2)c(=O)oc31	677	677.8079	16.76	16.314969	0.75	0.7380489	8.54	8.204046
46	CC1(C)CCN2CCC(C)(C)c3c2c1cc1cc(-c2ccc(s2)-c2ccc(\C=C(\C#N)C(O)=O)s2)c(=O)oc31	740	666.2336	16.56	16.331848	0.73	0.7420272	8.95	8.133166
47	CC1(C)CCN2CCC(C)(C)c3c2c1cc1cc(-c2ccc(s2)-c2ccc(\C=C(\C#N)C(O)=O)s2)c(=O)oc31	715	671.911	16.92	16.231744	0.73	0.7426824	8.88	8.152558
48	CC1(C)CCN2CCC(C)(C)c3c2c1cc1cc(-c2ccc(s2)-c2ccc(\C=C(\C#N)C(O)=O)s2)c(=O)oc31	702	674.1847	15.38	16.560493	0.78	0.7320497	8.47	8.281813
49	CC1(C)CCN2CCC(C)(C)c3c2c1cc1cc(-c2ccc(s2)-c2ccc(\C=C(\C#N)C(O)=O)s2)c(=O)oc31	663	681.1441	18.01	16.037023	0.76	0.7361694	9.09	8.089648
58	CC1(C)CCN2CCC(C)(C)c3c2c1cc1cc(-c2ccc(s2)-c2ccc(\C=C(\C#N)C(O)=O)s2)c(=O)oc31	581	698.9193	14.66	16.671624	0.69	0.7500053	5.97	8.768657
74	CC1(C)CCN2CCC(C)(C)c3c2c1cc1cc(\C=C(/C#N)c2ccc(\C=C(/C#N)C(O)=O)s2)c(=O)oc31	530	634.6897	10.93	12.630977	0.75	0.7270662	4.3	6.25443
76	CC1(C)CCN2CCC(C)(C)c3c2c1cc1cc(\C=C(/C#N)c2ccc(s2)-c2ccc(\C=C(\C#N)C(O)=O)s2)c(=O)oc31	580	609.8689	16.9	12.888639	0.74	0.7127731	7.3	5.687211
78	CC1(C)CC(\C=C\c2cc3cc4c5N(CCC4(C)C)CCC(C)(C)c5c3oc2=O)=C/C(/C1)=C(/C#N)C(O)=O	610	648.9378	14.6	12.124133	0.7	0.7399806	6.2	5.816908
79	CC1(C)CC(\C=C\c2cc3cc4c5N(CCC4(C)C)CCC(C)(C)c5c3oc2=O)=C/C(/C1)=C(/C#N)C(O)=O	620	646.7411	14.1	12.116595	0.74	0.7319502	6.5	5.724248

Molecule ID	Smiles	V_{OC}	Previsão V_{OC}	J_{SC}	Previsão J_{SC}	FF	Previsão FF	PCE	Previsão PCE
80	CC1(C)CC(\C=C\c2cc3cc4c5N(CCC4(C)C)CCC(C)(C)c5c3oc2=O)=C/C(/C1)=C(/C#N)C(O)=O	640	644.037	12.5	12.609094	0.73	0.7334897	5.9	5.891309
81	CC1(C)CC(\C=C\c2cc3cc4c5N(CCC4(C)C)CCC(C)(C)c5c3oc2=O)=C/C(/C1)=C(/C#N)C(O)=O	660	637.0629	11.2	12.952934	0.74	0.7306836	5.5	5.981371
82	CC1(C)CC(\C=C\c2cc3cc4c5N(CCC4(C)C)CCC(C)(C)c5c3oc2=O)=C/C(/C1)=C(/C#N)C(O)=O	680	631.6278	10.3	13.121205	0.75	0.728353	5.3	6.005511
88	CC1(C)CCN2CCC(C)(C)c3c2c1cc1cc(-c2ccc(\C=C\c4ccc(\C=C(\C#N)C(O)=O)s4)s2)c(=O)oc31	610	660.6817	13.3	11.56218	0.68	0.737648	5.5	5.261352
232	CCN1c2ccccc2Sc2cc(\C=C\c3sc(\C=C(/C#N)C(O)=O)c4OCCOc34)ccc12	645	632.7877	15.18	10.965269	0.69	0.6743363	6.72	4.760286
254	CCCCN1[C@@H]2CCCC[C@@H]2S[C@@H]2C[C@@H](CC[C@H]12)[C@H]1CC[C@@H]2[C@H]3CC[C@@H](C[C@@H]3C3(OCCO3)C3(OCCO3)[C@@H]2C1)[C@H]1CC[C@H](S1)\C=C(/C#N)C(O)=O	720	637.1003	11.1	11.836432	0.75	0.6950734	6	5.287234
359	CC(C)CCOc1ccc(cc1)N1c2ccccc2Sc2cc(\C=C\c3nc4ccc(cc4nc3\C=C\c3ccc4N(c5ccc(OCCC(C)C)cc5)c5ccccc5Sc4c3)C(O)=O)ccc12	610	659.8221	9.99	9.152798	0.71	0.6989024	4.36	4.349876
411	CCCCOc1ccc(c(OCCCC)c1)-c1ccc2N(CCCC)c3ccc(cc3Sc2c1)-c1ccc(-c2ccc(\C=C(/C#N)C(O)=O)s2)c2nsnc12	645	640.8076	14.71	12.959056	0.67	0.6628117	6.4	5.401524
440	CCCCN1c2ccc(cc2Sc2cc(ccc12)-c1nc2ccccc2n1-c1ccccc1)-c1ccc(\C=C(/C#N)C(O)=O)s1	606	628.8727	9.09	10.346904	0.66	0.6635816	3.66	4.33451
441	CCCCN1c2ccc(cc2Sc2cc(ccc12)-c1nc2ccccc2n1-c1ccccc1)-c1ccc(s1)-c1ccc(\C=C(/C#N)C(O)=O)s1	604	614.2149	9.26	10.94915	0.64	0.6633175	3.6	4.300175
442	CCCCN1c2ccc(cc2Sc2cc(ccc12)-c1ccc(\C=C(\C#N)C(O)=O)cc1)-c1nc2ccccc2n1-c1ccccc1	646	636.0112	7.47	9.387483	0.69	0.6724433	3.31	3.964449

Molecule ID	Smiles	V_{OC}	Previsão V_{OC}	J_{SC}	Previsão J_{SC}	FF	Previsão FF	PCE	Previsão PCE
443	CCCCN1c2ccc(cc2Sc2cc(ccc12)-c1nc2ccccc2n1-c1ccccc1)-c1ccc(s1)-c1ccc(\C=C(\C#N)C(O)=O)cc1	630	628.2228	7.53	9.206144	0.68	0.6692297	3.24	4.201459
444	CCCCN1c2ccccc2Sc2cc(ccc12)-c1ccc(s1)-c1ccc(\C=C(\C#N)C(O)=O)s1	519	635.6793	5.46	11.312053	0.58	0.677797	1.63	4.994671
455	CCCCN1c2ccccc2Sc2cc(\C=C\c3ccc(\C=C(/C#N)C(O)=O)s3)ccc12	685	620.585	13.4	10.902871	0.7	0.6953742	6.4	4.691744
456	CCCCN1c2ccccc2Sc2cc(\C=C\c3ccc(\C=C(/C#N)C(O)=O)s3)ccc12	638	649.6408	12.2	11.961619	0.72	0.6854329	5.6	5.157694
461	CCn1c2ccccc2c2cc(ccc12)N(c1ccc2-c3ccccc3C(CC)(CC)c2c1)c1ccc2-c3ccc(cc3C(CC)(CC)c2c1)-c1ccc(s1)-c1ccc(\C=C(/C#N)C(O)=O)s1	571	626.6788	6.86	10.936065	0.69	0.6337197	2.71	4.573553
491	CCCCCCN1c2ccc(C=C3SC(=S)N(CC(O)=O)C3=O)cc2Sc2cc(ccc12)-n1c2ccc(cc2c2cc(ccc12)-n1c2ccc(cc2c2cc(ccc12)C(C)(C)C)C(C)(C)C)-n1c2ccc(cc2c2cc(ccc12)C(C)(C)C)C(C)(C)C	658	675.5022	7.59	8.95738	0.63	0.6586218	3.31	3.603856
492	CCCCN1c2ccc(C=C3SC(=S)N(CC(O)=O)C3=O)cc2Sc2cc(ccc12)-n1c2ccc(cc2c2cc(ccc12)-n1c2ccc(cc2c2cc(ccc12)C(C)(C)C)C(C)(C)C)-n1c2ccc(cc2c2cc(ccc12)C(C)(C)C)C(C)(C)C	664	649.0902	7.07	9.089648	0.66	0.6438637	3.1	3.649354
493	CCCCCCN1c2ccc(C=C3SC(=S)N(CC(O)=O)C3=O)cc2Sc2cc(ccc12)-n1c2ccc(cc2c2cc(ccc12)C(C)(C)C)C(C)(C)C	660	669.6093	9.06	9.414989	0.63	0.6707852	3.58	4.264159
495	CCCCCCCCc1cc(\C=C(/C#N)C(O)=O)sc1-c1ccc(s1)-c1sc(cc1CCCCCCCC)-c1cc2cc(-c3ccc(OCCCCCC)cc3)n3c2c(c1)sc1ccccc31	726	700.3762	17.76	13.119469	0.58	0.692096	7.54	6.913306

Molecule ID	Smiles	V_{OC}	Previsão V_{OC}	J_{SC}	Previsão J_{SC}	FF	Previsão FF	PCE	Previsão PCE
497	CCn1c2ccc(cc2c2ccc(cc12)-c1ccc(\C=C(/C#N)C(O)=O)s1)N(c1ccc2-c3ccc(cc3C(CC)(CC)c2c1)-c1ccc(\C=C(/C#N)C(O)=O)s1)c1ccc2-c3ccc(cc3C(CC)(CC)c2c1)-c1ccc(\C=C(/C#N)C(O)=O)s1	497	598.8125	2.04	7.818392	0.68	0.6540567	0.69	3.995856
515	CCCCCCOc1ccc(cc1)-c1cc2cc(cc3sc4ccccc4n1c23)-c1ccc(\C=C(/C#N)C(O)=O)o1	705	687.739	12.6	11.065876	0.63	0.6640985	5.6	5.183276
516	CCCCCCOc1ccc(cc1)-c1cc2cc(cc3sc4ccccc4n1c23)-c1ccc(\C=C(/C#N)C(O)=O)s1	736	665.8351	13.38	10.489759	0.65	0.6520339	6.42	4.825412
543	CCCCCCN1c2ccc(\C=C(\C#N)C(O)=O)cc2Sc2cc(ccc12)-n1c2ccc(cc2c2cc(ccc12)-n1c2ccc(cc2c2cc(ccc12)C(C)(C)C)C(C)(C)C)-n1c2ccc(cc2c2cc(ccc12)C(C)(C)C)C(C)(C)C	745	704.1734	13.11	12.057075	0.66	0.6662851	6.17	5.805162
544	CCCCN1c2ccc(\C=C(\C#N)C(O)=O)cc2Sc2cc(ccc12)-n1c2ccc(cc2c2cc(ccc12)-n1c2ccc(cc2c2cc(ccc12)C(C)(C)C)C(C)(C)C)-n1c2ccc(cc2c2cc(ccc12)C(C)(C)C)C(C)(C)C	739	673.2394	14.68	11.626989	0.68	0.6596737	7.11	5.508765
545	CCCCCCN1c2ccc(\C=C(\C#N)C(O)=O)cc2Sc2cc(ccc12)-n1c2ccc(cc2c2cc(ccc12)C(C)(C)C)C(C)(C)C	745	711.8781	16.92	14.107637	0.67	0.6595227	7.97	6.546393
546	CCCCN1c2ccc(\C=C(\C#N)C(O)=O)cc2Sc2cc(ccc12)-n1c2ccc(cc2c2cc(ccc12)C(C)(C)C)C(C)(C)C	713	688.2637	15.99	13.53912	0.66	0.6679979	7.55	6.464175
547	CCCCCCCCCCCCN1c2ccc(\C=C(/C#N)C(O)=O)cc2Sc2cc(ccc12)-c1ccc(-c2ccc3N(CCCCCCCCCCCC)c4ccc(\C=C(/C#N)C(O)=O)cc4Sc3c2)c2nsnc12	728	727.4318	13.1	11.267642	0.72	0.6906408	6.87	5.531217
558	CCCCC(CC)CN1c2ccc(cc2Sc2cc(ccc12)-c1ccc2n(CC(CC)CCCC)c3nc4ccccc4nc3c2c1)-c1ccc(\C=C(/C#N)C(O)=O)o1	745	738.7928	14.2	11.659437	0.71	0.6853274	7.56	6.151378

Molecule ID	Smiles	V_{OC}	Previsão V_{OC}	J_{SC}	Previsão J_{SC}	FF	Previsão FF	PCE	Previsão PCE
559	CCCCC(CC)CN1c2ccc(cc2Sc2cc(ccc12)-c1ccc2n(CC(CC)CCCC)c3nc4ccccc4nc3c2c1)-c1ccc(\C=C(/C#N)C(O)=O)s1	757	714.7941	15.3	12.432233	0.71	0.6744571	8.28	6.107073
560	CCCCC(CC)CN1c2ccccc2Sc2cc(ccc12)-c1ccc(\C=C(/C#N)C(O)=O)s1	705	682.6162	14	9.968538	0.69	0.6709573	6.82	5.54298
588	CCN1c2ccc(C=C3SC(=S)N(CC(O)=O)C3=O)cc2Sc2cc(C=C3SC(=S)N(CC(O)=O)C3=O)ccc12	658	602.511	10.6	9.013582	0.7	0.665188	4.91	3.805065
637	CCn1c2ccc(cc2c2ccc(cc12)-c1ccc(s1)-c1ccc(\C=C(/C#N)C(O)=O)s1)N(c1ccc2-c3ccc(cc3C(CC)(CC)c2c1)-c1ccc(s1)-c1ccc(\C=C(/C#N)C(O)=O)s1)c1ccc2-c3ccc(cc3C(CC)(CC)c2c1)-c1ccc(s1)-c1ccc(\C=C(/C#N)C(O)=O)s1	512	567.0089	3.72	7.97333	0.63	0.6493167	1.21	3.672433
640	CCn1c2ccccc2c2cc(ccc12)N(c1ccc2-c3ccccc3C(CC)(CC)c2c1)c1ccc2-c3ccc(cc3C(CC)(CC)c2c1)-c1ccc(\C=C(/C#N)C(O)=O)s1	650	600.4667	8.3	10.201783	0.69	0.6405814	3.73	3.97642
653	OC(=O)C(=Cc1ccc(cc1)-n1c2ccccc2c2ccccc12)C#N	740	584.3491	4.42	7.330883	0.77	0.7070948	2.36	3.18703
669	CCCCn1c2cc(ccc2c2ccc(cc12)-c1ccc(s1)-c1ccc(\C=C(/C#N)C(O)=O)s1)N(c1ccccc1)c1ccccc1	630	637.223	18.1	13.871924	0.6	0.6394737	6.83	5.561035
670	CCCCn1c2cc(ccc2c2ccc(cc12)-c1ccc(\C=C(/C#N)C(O)=O)s1)N(c1ccccc1)c1ccccc1	660	660.8008	15.3	12.404806	0.6	0.6433367	6.05	5.300901
686	CCCCCCn1c2ccc(C=C3SC(=S)N(CC(O)=O)C3=O)cc2c2cc(C=C3SC(=S)N(CC(O)=O)C3=O)ccc12	589	647.2805	7.61	8.48225	0.63	0.6633211	2.81	3.866765
724	CCCCCCCCCCCCn1c2ccc(\C=C(/C#N)C(O)=O)cc2c2cc(ccc12)-c1ccc(-c2ccc3c(c2)n(CCCCCCCCCCCC)c2ccc(\C=C(/C#N)C(O)=O)cc32)c2nsnc12	762	722.6189	7.51	11.975493	0.76	0.6861897	4.35	6.251643

Molecule ID	Smiles	V_{OC}	Previsão V_{OC}	J_{SC}	Previsão J_{SC}	FF	Previsão FF	PCE	Previsão PCE
765	CCCCOc1ccc(cc1)N(c1ccc(OCCCC)cc1)c1ccc2c(c1)n(CCCC)c1c c(ccc21)-c1ccc(s1)-c1ccc(\C=C(/C#N)C(O)=O)s1	634	665.5855	14.6	12.913874	0.62	0.6534812	5.76	5.87548
767	CCCCn1c2cc(ccc2c2ccc(cc12)-c1ccc(s1)- c1ccc(\C=C(/C#N)C(O)=O)s1)N(c1ccccc1)c1ccccc1	644	626.9909	14.5	15.823396	0.63	0.6222466	5.84	6.166067
768	CCCCn1c2cc(ccc2c2ccc(cc12)- c1ccc(\C=C(/C#N)C(O)=O)s1)N(c1ccccc1)c1ccccc1	654	664.901	12.4	14.25874	0.64	0.6206304	5.16	5.867949
772	CCCCCCCCn1c2ccccc2c2c(- c3ccc(OCC)cc3)c3n(CCCCCCCC)c4ccc(cc4c3c(- c3ccc(OCC)cc3)c12)-c1ccc(\C=C(/C#N)C(O)=O)s1	738	728.6072	13.9	13.466844	0.68	0.6820706	6.98	6.518997
773	CCCCCCCCn1c2ccccc2c2c(- c3ccc(cc3)N(CC)CC)c3n(CCCCCCCC)c4ccc(cc4c3c(- c3ccc(cc3)N(CC)CC)c12)-c1ccc(\C=C(/C#N)C(O)=O)s1	745	730.3653	15.2	13.022308	0.71	0.6697423	8.09	6.158009
774	CCCCCCCCn1c2ccccc2c2c(- c3ccc(CC)cc3)c3n(CCCCCCCC)c4ccc(cc4c3c(- c3ccc(CC)cc3)c12)-c1ccc(\C=C(/C#N)C(O)=O)s1	729	741.3261	12.6	13.471344	0.68	0.6768077	6.25	6.786114
775	CCCCCCCCc1cc(C=C(C#N)C(O)=O)sc1-c1ccc(s1)- c1sc(cc1CCCCCCCC)- c1cc2c3cc(OC)ccc3n(C(CC)CCCCC)c2c2ccccc12	744	715.6158	14.8	14.212237	0.68	0.6794561	7.54	6.958902
776	CCCCCC(CC)n1c2ccc(OC)cc2c2cc(- c3ccc(C=C(C#N)C(O)=O)o3)c3ccccc3c12	757	713.8759	13.8	12.631386	0.66	0.6669142	6.93	5.790548
777	CCCCCC(CC)n1c2ccc(OC)cc2c2cc(- c3ccc(C=C(C#N)C(O)=O)s3)c3ccccc3c12	729	718.5436	12.4	12.658534	0.66	0.6718627	6.01	5.853924
814	CCCCn1c2cc(ccc2c2ccc(cc12)-c1ccc(s1)- c1ccc(\C=C(/C#N)C(O)=O)s1)N(c1ccc(cc1)-c1ccc2- c3ccccc3C(CC)(CC)c2c1)c1ccc(cc1)-c1ccc2- c3ccccc3C(CC)(CC)c2c1	596	592.1926	11.5	9.820485	0.61	0.6181935	4.2	3.619239

Molecule ID	Smiles	V_{OC}	Previsão V_{OC}	J_{SC}	Previsão J_{SC}	FF	Previsão FF	PCE	Previsão PCE
815	CCCCn1c2cc(ccc2c2ccc(cc12)-c1ccc(\C=C(/C#N)C(O)=O)s1)N(c1ccc(cc1)-c1ccc2-c3ccccc3C(CC)(CC)c2c1)c1ccc(cc1)-c1ccc2-c3ccccc3C(CC)(CC)c2c1	536	626.2571	6.02	11.564998	0.53	0.6632417	1.78	4.458593
918	CCCCn1c(nc2c1c1cccc3ccc4cccc2c4c13)-c1ccc(cc1)-c1ccc(\C=C(/C#N)C(O)=O)s1	545	586.8249	8.8	9.633264	0.71	0.6662805	3.41	4.280294
956	C1CCN(CC1)c1ccc(\C=C\c2ccncc2)c2nsnc12	516	576.0781	1.99	4.861916	0.61	0.6505628	0.62	2.069313
957	C1CN(CCO1)c1ccc(\C=C\c2ccncc2)c2nsnc12	552	543.2224	2.42	5.048449	0.66	0.6282304	0.89	2.083207
958	C(=C/c1ccc(\C=C\c2ccncc2)c2nsnc12)\c1ccncc1	445	593.059	0.56	5.359544	0.58	0.6390053	0.14	2.10647
959	CCCCn1c2ccccc2c2cc(\C=C\c3ccc(\C=C\c4ccncc4)c4nsnc34)ccc12	516	578.9939	3.96	6.263325	0.64	0.6416133	1.31	2.550562
960	CCCCN1c2ccccc2Sc2cc(\C=C\c3ccc(\C=C\c4ccncc4)c4nsnc34)ccc12	564	573.5139	5.65	6.526667	0.62	0.6486691	1.97	2.856631
1041	CCCCC(CC)Cn1c2cc(ccc2c2nc3ccccc3nc12)-c1ccc(\C=C(/C#N)C(O)=O)s1	676	679.9483	8.9	10.164853	0.68	0.654072	4.1	4.816493
1042	CCn1c(nc(c1-c1ccccc1)-c1ccccc1)-c1ccc2-c3ccc(cc3C(CC)(CC)c2c1)-c1ccc(\C=C(/C#N)C(O)=O)s1	610	596.6854	5.78	8.724841	0.68	0.6619617	2.42	3.505994
1043	CCn1c(nc(c1-c1ccccc1)-c1ccccc1)-c1ccc2-c3ccc(cc3C(CC)(CC)c2c1)-c1ccc(s1)-c1ccc(\C=C(/C#N)C(O)=O)s1	570	600.9776	9.25	9.369563	0.66	0.6592392	3.44	3.278256
1044	CCn1c(nc(c1-c1ccccc1)-c1ccccc1)-c1cc2c(-c3ccc(cc3C2(CC)CC)-c2ccc(\C=C(/C#N)C(O)=O)s2)c(c1)-c1nc(c(-c2ccccc2)n1CC)-c1ccccc1	600	601.2678	4.16	8.253412	0.68	0.6680974	1.67	3.21709
1045	CCn1c(nc(c1-c1ccccc1)-c1ccccc1)-c1cc2c(-c3ccc(cc3C2(CC)CC)-c2ccc(s2)-c2ccc(\C=C(/C#N)C(O)=O)s2)c(c1)-c1nc(c(-c2ccccc2)n1CC)-c1ccccc1	580	600.9321	7.13	8.762063	0.67	0.6622605	2.75	3.056571

Molecule ID	Smiles	V_{OC}	Previsão V_{OC}	J_{SC}	Previsão J_{SC}	FF	Previsão FF	PCE	Previsão PCE
1046	CCn1c(nc(c1-c1ccc2-c3ccccc3C(CC)(CC)c2c1)-c1ccc2-c3ccccc3C(CC)(CC)c2c1)-c1cc2c(-c3ccc(cc3C2(CC)CC)-c2ccc(\C=C(/C#N)C(O)=O)s2)c(c1)-c1nc(c(-c2ccc3-c4ccccc4C(CC)(CC)c3c2)n1CC)-c1ccc2-c3ccccc3C(CC)(CC)c2c1	570	617.4021	3.76	8.11398	0.65	0.6558344	1.39	3.979549
1047	CCn1c(nc(c1-c1ccc2-c3ccccc3C(CC)(CC)c2c1)-c1ccc2-c3ccccc3C(CC)(CC)c2c1)-c1cc2c(-c3ccc(cc3C2(CC)CC)-c2ccc(s2)-c2ccc(\C=C(/C#N)C(O)=O)s2)c(c1)-c1nc(c(-c2ccc3-c4ccccc4C(CC)(CC)c3c2)n1CC)-c1ccc2-c3ccccc3C(CC)(CC)c2c1	630	584.7523	7.88	7.397899	0.66	0.649433	3.27	3.128519
1050	CC(C)(C)c1ccc(cc1)-c1sc(-c2ccc(cc2)C(C)(C)C)c2[nH]c(nc12)-c1ccc(cc1)-c1ccc(\C=C(/C#N)C(O)=O)s1	566	619.43	1.06	9.379102	0.57	0.6264757	0.34	3.656612
1051	CC(C)(C)c1ccc(cc1)-c1sc(-c2ccc(cc2)C(C)(C)C)c2[nH]c(nc12)-c1ccc(s1)-c1ccc(\C=C(/C#N)C(O)=O)s1	601	597.7203	6.93	9.33306	0.64	0.5995906	2.67	3.394009
1060	CCCCn1c(nc2c1c1cccc3ccc4cccc2c4c13)-c1ccc(s1)-c1ccc(\C=C(/C#N)C(O)=O)s1	524	575.747	11.4	10.798479	0.63	0.6778062	3.75	4.254734
1061	CCCCn1c(nc2c1c1cccc3ccc4cccc2c4c13)-c1ccc(s1)-c1ccc(s1)-c1ccc(\C=C(/C#N)C(O)=O)s1	543	576.9975	15.5	10.694711	0.67	0.6480196	5.65	3.758641
1086	CC(C)(C)c1ccc(cc1)-c1sc(-c2ccc(cc2)C(C)(C)C)c2[nH]c(nc12)-c1ccc(cc1)-c1ccc(\C=C(/C#N)C(O)=O)cc1	583	617.7708	3.09	6.906569	0.56	0.6321952	1.02	2.56835
1087	CCCCn1c(nc2c(sc(-c3cccs3)c12)-c1cccs1)-c1ccc(cc1)-c1ccc(\C=C(/C#N)C(O)=O)cc1	573	581.479	4.07	8.326438	0.69	0.6394761	1.6	3.244001
1089	CCCCn1c(nc2c(sc(-c3cccs3)c12)-c1cccs1)-c1ccc(s1)-c1ccc(\C=C(/C#N)C(O)=O)s1	558	590.5572	7.91	9.530394	0.63	0.6485604	2.78	4.059814
1095	CCCCCCOc1ccc(c(OCCCCCC)c1)-n1c2ccsc2c2sc(\C=C(/C#N)C(O)=O)cc12	560	711.5862	7.5	12.664098	0.7	0.6551929	2.94	6.406524

Molecule ID	Smiles	V_{OC}	Previsão V_{OC}	J_{SC}	Previsão J_{SC}	FF	Previsão FF	PCE	Previsão PCE
1181	CCCCCCCCOc1cc2c3cc(CC(CN)C(O)O)sc3c3sc(cc3c2cc1OCCCCCCCC)-c1ccc(-c2ccc(cc2)N(c2ccc(OC)cc2)c2ccc(OC)cc2)c2nsnc12	658	681.093	15.84	8.94762	0.68	0.6691179	7.1	4.470766
1240	CCN1c2ccc(C=C3SC(=S)N(CC(O)=O)C3=O)cc2CCc2cc(C=C3SC(=S)N(CC(O)=O)C3=O)ccc12	597	618.2301	9.95	8.804639	0.61	0.6636975	3.59	3.952676
1269	CCCCCCc1nc(sc1-c1ccc(\C=C(/C#N)C(O)=O)s1)-c1nc(CCCCCC)c(s1)-c1ccc(cc1)N(c1ccccc1)c1ccccc1	810	727.7652	11.78	11.505599	0.6	0.6299486	5.73	5.149134
1270	CCCCCCc1nc(sc1C#Cc1ccc(cc1)N(c1ccccc1)c1ccccc1)-c1nc(CCCCCC)c(s1)-c1ccc(\C=C(/C#N)C(O)=O)s1	754	729.6236	12.06	12.022992	0.62	0.6188183	5.6	5.241169
1283	CCCCCCOc1ccc(cc1)N(c1ccc(OCCCCCC)cc1)c1ccc(cc1)-c1ccc(s1)-c1cc(cs1)-c1nc2cccc(-c3ccc(s3)-c3ccc(cc3)N(c3ccc(OCCCCCC)cc3)c3ccc(OCCCCCC)cc3)c2nc1-c1ccc(\C=C(/C#N)C(O)=O)s1	681	715.5434	12	12.232259	0.62	0.6494037	5.1	5.648259
1284	CCCCCCOc1ccc(cc1)N(c1ccc(OCCCCCC)cc1)c1ccc(cc1)-c1ccc(s1)-c1ccc(-c2ccc(s2)-c2ccc(cc2)N(c2ccc(OCCCCCC)cc2)c2ccc(OCCCCCC)cc2)c2nc(-c3ccc(\C=C(\C#N)C(O)=O)s3)c(nc12)-c1ccc(\C=C(\C#N)C(O)=O)s1	725	687.752	11.6	10.05591	0.63	0.6501991	5.2	4.309915
1285	CCCCCCOc1ccc(cc1)N(c1ccc(OCCCCCC)cc1)c1ccc(cc1)-c1ccc(s1)-c1cc(cs1)-c1nc2cccc(-c3ccc(s3)-c3ccc(cc3)N(c3ccc(OCCCCCC)cc3)c3ccc(OCCCCCC)cc3)c2nc1-c1ccc(\C=C(/C#N)C(O)=O)s1	720	690.6498	13.2	11.534635	0.65	0.6307685	6.1	5.00565
1292	CCCCCCc1nc(sc1-c1ccc(\C=C(/C#N)C(O)=O)s1)-c1nc(CCCCCC)c(s1)-c1ccc(cc1)N(c1ccc(OC)cc1)c1ccc(OC)cc1	745	717.8467	10.22	11.269384	0.59	0.6326748	4.48	5.708119
1293	CCCCCCc1nc(sc1-c1ccc(\C=C(/C#N)C(O)=O)cc1)-c1nc(CCCCCC)c(s1)-c1ccc(cc1)N(c1ccc(OC)cc1)c1ccc(OC)cc1	782	727.0761	8.92	10.856302	0.6	0.6359673	4.2	5.262464

Molecule ID	Smiles	V_{OC}	Previsão V_{OC}	J_{SC}	Previsão J_{SC}	FF	Previsão FF	PCE	Previsão PCE
1294	CCCCCCc1nc(sc1C#Cc1ccc(\C=C(/C#N)C(O)=O)cc1)-c1nc(CCCCCC)c(s1)-c1ccc(cc1)N(c1ccc(OC)cc1)c1ccc(OC)cc1	790	725.6704	10.51	10.547678	0.61	0.6217905	5.05	5.100665
1312	OC(=O)C(=C\c1ccc(s1)-c1ccc(-c2ccc(s2)-c2ccc(cc2)N(c2ccccc2)c2ccccc2)c2nsnc12)\C#N	540	630.0872	11.9	10.811752	0.59	0.6680117	3.81	4.480155
1381	COc1ccc(cc1)N1c2ccccc2Sc2cc(\C=C\c3nc4ccc(cc4nc3\C=C\c3ccc4N(c5ccc(OC)cc5)c5ccccc5Sc4c3)C(O)=O)ccc12	610	638.8751	9.99	9.477718	0.71	0.6699006	4.36	4.299708
1382	COc1ccc(cc1)N1c2ccccc2Sc2cc(\C=C\c3nc4ccc(cc4nc3\C=C\c3ccc4N(c5ccc(OC)cc5)c5ccccc5Sc4c3)C(O)=O)ccc12	640	619.7235	9.77	9.43711	0.67	0.6962803	4.18	4.399139
1392	CCCCCCOc1ccc(cc1)N(c1ccc(OCCCCCC)cc1)c1ccc(cc1)-c1ccc(s1)-c1ccc(-c2ccc(s2)-c2ccc(cc2)N(c2ccc(OCCCCCC)cc2)c2ccc(OCCCCCC)cc2)c2nc(-c3ccc(\C=C(\C#N)C(O)=O)s3)c(nc12)-c1ccc(\C=C(\C#N)C(O)=O)s1	682	715.7864	9.1	11.630054	0.65	0.6385932	4	5.112672
1393	CC(C)CCOc1ccc(cc1)N(c1ccc(OCCC(C)C)cc1)c1ccc(\C=C\c2nc3ccc(cc3nc2\C=C\c2ccc(cc2)N(c2ccc(OCCC(C)C)cc2)c2ccc(OCCC(C)C)cc2)C(O)=O)cc1	700	654.8593	7.6	9.446797	0.75	0.6777495	3.98	4.282353
1442	CC(C)(C)c1ccc(cc1)-c1ccc(cc1)N(c1ccc(\C=C\c2ccc(\C=C(\C#N)C(O)=O)o2)cc1)c1ccc(cc1)-c1ccc(cc1)C(C)(C)C	720	660.2691	10.98	9.974595	0.57	0.6439203	4.48	4.590321
1443	CC(C)(C)c1ccc(cc1)-c1ccc(cc1)N(c1ccc(\C=C\c2sc(\C=C(\C#N)C(O)=O)c3OCCOc23)cc1)c1ccc(cc1)-c1ccc(cc1)C(C)(C)C	640	656.1466	10.31	12.144502	0.62	0.6348777	4.08	5.367084
1444	CC(C)(C)c1ccc(cc1)-c1ccc(cc1)N(c1ccc(\C=C\c2ccc(s2)-c2ccc(\C=C(\C#N)C(O)=O)s2)cc1)c1ccc(cc1)-c1ccc(cc1)C(C)(C)C	615	652.0605	11.8	11.791652	0.58	0.6278822	4.23	4.92394

Molecule ID	Smiles	V_{OC}	Previsão V_{OC}	J_{SC}	Previsão J_{SC}	FF	Previsão FF	PCE	Previsão PCE
1445	CC(C)(C)c1ccc(cc1)-c1ccc(cc1)N(c1ccc(\C=C\c2ccc(\C=C(\C#N)C(O)=O)s2)cc1)c1ccc(cc1)-c1ccc(cc1)C(C)(C)C	675	650.0317	12.65	11.295888	0.63	0.6028016	5.35	4.448557
1510	CCC1(CC)c2cc(ccc2-c2ccc(cc12)-c1ccc(\C=C(/C#N)C(O)=O)s1)N(c1ccccc1)c1ccc(C=C2c3ccccc3-c3ccccc23)cc1	642	651.4653	14.1	11.898894	0.66	0.6521322	5.98	5.049637
1511	CCC1(CC)c2cc(ccc2-c2ccc(cc12)-c1ccc(s1)-c1ccc(\C=C(/C#N)C(O)=O)s1)N(c1ccccc1)c1ccc(C=C2c3ccccc3-c3ccccc23)cc1	650	635.5135	14.8	11.867247	0.64	0.6474103	6.13	5.083434
1512	CCC1(CC)c2cc(ccc2-c2ccc(cc12)-c1ccc(\C=C(/C#N)C(O)=O)s1)N(c1ccc(C=C2c3ccccc3-c3ccccc23)cc1)c1ccc(C=C2c3ccccc3-c3ccccc23)cc1	650	645.1185	13.7	12.59254	0.65	0.646075	5.82	5.405368
1513	CCC1(CC)c2cc(ccc2-c2ccc(cc12)-c1ccc(s1)-c1ccc(\C=C(/C#N)C(O)=O)s1)N(c1ccc(C=C2c3ccccc3-c3ccccc23)cc1)c1ccc(C=C2c3ccccc3-c3ccccc23)cc1	640	633.3631	13.2	13.039864	0.64	0.6417413	5.38	5.28449
1514	CCC1(CC)c2cc(ccc2-c2ccc(cc12)-c1ccc(\C=C(/C#N)C(O)=O)s1)N(c1ccccc1)c1ccc(C=C(C#N)C#N)cc1	586	607.9386	10.3	11.968551	0.67	0.6534058	4.01	5.065655
1515	CCC1(CC)c2cc(ccc2-c2ccc(cc12)-c1ccc(\C=C(/C#N)C(O)=O)s1)N(c1ccccc1)c1ccccc1	690	638.6152	10.56	10.88595	0.65	0.6611712	4.7	4.157553
1516	CCC1(CC)c2cc(ccc2-c2ccc(cc12)-c1ccc(s1)-c1ccc(\C=C(/C#N)C(O)=O)s1)N(c1ccccc1)c1ccccc1	618	642.3261	7.52	12.792418	0.65	0.6548128	3.01	4.880329
1525	CCCCCCc1cc(\C=C(/C#N)C(O)=O)sc1-c1cc(CCCCCC)c(s1)-c1cc(CCCCCC)c(s1)-c1ccc(cc1)N(c1ccc2-c3ccccc3C(C)(C)c2c1)c1ccc2-c3ccccc3C(C)(C)c2c1	641	690.18	16.13	13.666298	0.72	0.6670951	7.42	6.17789

Molecule ID	Smiles	V_{OC}	Previsão V_{OC}	J_{SC}	Previsão J_{SC}	FF	Previsão FF	PCE	Previsão PCE
1526	CCCCCCC1=CC(SC1c1cc2ccc(cc2s1)N(c1ccc2-c3ccccc3C(C)(C)c2c1)c1ccc2-c3ccccc3C(C)(C)c2c1)=c1sc(cc1CCCCCC)=C1SC(C=O)(\C=C(/C#N)C(O)=O)C=C1CCCCCC	664	667.0646	17.45	12.202688	0.74	0.6673909	8.6	5.901853
1822	CCCCCCOc1ccc(cc1)N(c1ccc(OCCCCCC)cc1)c1ccc(cc1)-c1ccc2c(-c3ccc(OCCCCCC)cc3)c3cc(ccc3c(-c3ccc(OCCCCCC)cc3)c2c1)-c1ccc(s1)-c1ccc(\C=C(\C#N)C(O)=O)s1	695	711.2676	13.95	12.543588	0.65	0.6827907	6.34	6.076354
1823	CCCCCCOc1ccc(cc1)N(c1ccc(OCCCCCC)cc1)c1ccc(cc1)-c1ccc(s1)-c1ccc2c(-c3ccc(OCCCCCC)cc3)c3cc(ccc3c(-c3ccc(OCCCCCC)cc3)c2c1)-c1ccc(\C=C(\C#N)C(O)=O)s1	722	700.0694	12.95	12.476996	0.69	0.6649744	6.44	5.719131
1824	CCCCCCOc1ccc(cc1)N(c1ccc(OCCCCCC)cc1)c1ccc(cc1)-c1ccc(s1)-c1ccc(s1)-c1ccc2c(-c3ccc(OCCCCCC)cc3)c3cc(C=C(C#N)C(O)=O)ccc3c(-c3ccc(OCCCCCC)cc3)c2c1	698	708.8741	10.68	12.440664	0.69	0.6707524	5.15	6.038983
1837	OC(=O)C[n+]1ccc(\C=C\c2ccc(cc2)N(c2ccccc2)c2ccccc2)cc1	614	661.0946	5.9	7.69387	0.73	0.705703	2.6	3.687553
1856	OC(=O)C(=Cc1ccc(s1)-c1c2ccccc2c(-c2ccc(cc2)N(c2ccccc2)c2ccccc2)c2ccccc12)C#N	708	657.758	5.54	10.597721	0.79	0.7333876	3.09	5.181444
1857	OC(=O)C(=Cc1ccc(C=Cc2c3ccccc3c(-c3ccc(cc3)N(c3ccccc3)c3ccccc3)c3ccccc23)cc1)C#N	726	676.6367	8.78	8.991904	0.81	0.7348475	5.14	4.661262
1918	OC(=O)C(=Cc1ccc(cc1)-c1c2ccccc2c(-c2ccc(cc2)N(c2ccccc2)c2ccccc2)c2ccccc12)C#N	766	673.1956	11.7	7.142582	0.76	0.7805721	6.78	3.486032
1919	OC(=O)C(=Cc1ccc(cc1)-c1c2ccccc2c(-c2ccc(cc2)N(c2ccccc2)c2ccccc2)c2ccccc12)C#N	678	728.804	5.39	10.832595	0.8	0.7610959	2.91	5.904172
1920	OC(=O)C(=Cc1ccc(C=Cc2c3ccccc3c(-c3ccc(cc3)N(c3ccccc3)c3ccccc3)c3ccccc23)s1)C#N	642	673.6178	10.6	11.007367	0.71	0.7301734	4.86	5.160699

Molecule ID	Smiles	V_{OC}	Previsão V_{OC}	J_{SC}	Previsão J_{SC}	FF	Previsão FF	PCE	Previsão PCE
1921	OC(=O)C(=C\c1ccc(s1)C#Cc1c2ccccc2c(-c2ccc(cc2)N(c2ccccc2)c2ccccc2)c2ccccc12)\C#N	608	667.8324	10.05	10.81451	0.74	0.718058	4.5	5.183874
1922	COc1ccc(cc1)N(c1ccc(OC)cc1)c1ccc(cc1)-c1c2ccccc2c(C#Cc2ccc(\C=C(/C#N)C(O)=O)cc2)c2ccccc12	720	640.9187	12.96	11.586646	0.75	0.7141407	7.03	5.424139
1923	COc1ccc(cc1)N(c1ccc(OC)cc1)c1ccc(cc1)-c1c2ccccc2c(C#Cc2ccc(\C=C(/C#N)C(O)=O)s2)c2ccccc12	677	646.0489	13.37	12.092864	0.75	0.6859344	6.82	5.9586
1924	OC(=O)C(=C\c1ccc(cc1)-c1c2ccccc2c(C#Cc2ccc(cc2)N(c2ccccc2)c2ccccc2)c2ccccc12)\C#N	648	690.4612	8.15	10.123123	0.74	0.7463475	3.93	4.969631
1925	COc1ccc(cc1)N(c1ccc(OC)cc1)c1ccc(cc1)C#Cc1c2ccccc2c(-c2ccc(\C=C(/C#N)C(O)=O)cc2)c2ccccc12	599	688.9567	13.07	11.64196	0.72	0.7173634	5.66	5.726132
1932	CCC1(CC)c2cc(ccc2-c2ccc(cc12)N(c1ccccc1)c1ccc(\C=C(/C#N)C(O)=O)cc1)N(c1ccccc1)c1ccccc1	571	648.9056	10.4	10.371905	0.69	0.6828396	4.11	4.78323
1933	CCC1(CC)c2cc(ccc2-c2ccc(cc12)N(c1ccc(\C=C(/C#N)C(O)=O)cc1)c1cccc2ccccc12)N(c1ccccc1)c1cccc2ccccc12	595	632.0616	10.5	10.413647	0.7	0.6747186	4.36	4.720892
1934	CCC1(CC)c2cc(ccc2-c2ccc(cc12)-n1c2ccccc2c2cc(\C=C(/C#N)C(O)=O)ccc12)-n1c2ccccc2c2ccccc12	605	625.7602	5.46	10.20477	0.7	0.6696072	2.3	4.271382
1935	CCC1(CC)c2ccccc2-c2ccc(cc12)N(c1ccccc1)c1ccc(\C=C(/C#N)C(O)=O)cc1	603	654.1532	8.73	10.189025	0.69	0.6824351	3.6	4.682158
1947	OC(=O)C[n+]1ccc(\C=C\c2ccc(cc2)N(c2ccccc2)c2ccccc2)cc1	666	624.8704	8.1	6.228872	0.73	0.7068243	3.9	2.907414
1975	CCCCCCOc1ccccc1-n1c2ccsc2c2sc(cc12)-c1ccc(cc1)N(c1ccc(cc1)-c1cc2n(-c3ccccc3OCCCCCC)c3cc(\C=C(/C#N)C(O)=O)sc3c2s1)c1cc2c(cc1OCCCCCC)n(CCCCCC)c1ccccc21	650	694.2726	12.3	11.768566	0.68	0.6604569	5.43	5.535513

Molecule ID	Smiles	V_{OC}	Previsão V_{OC}	J_{SC}	Previsão J_{SC}	FF	Previsão FF	PCE	Previsão PCE
2115	CCCCCCOc1ccc(c(OCCCCCC)c1)-n1c2cc(\C=C(/C#N)C(O)=O)sc2c2sc(cc12)-c1ccc(cc1)N(c1ccccc1)c1ccccc1	692	693.1178	16.3	12.533746	0.68	0.6616561	7.67	5.638562
2132	CCCCCCOc1ccc(c(OCCCCCC)c1)-n1c2cc(\C=C(/C#N)C(O)=O)sc2c2sc(cc12)-c1cc2n(-c3cc(OCCCCCC)cc(OCCCCCC)c3)c3cc(sc3c2s1)-c1ccc(cc1)N(c1ccccc1)c1ccccc1	582	686.5663	10.6	13.418769	0.68	0.6669004	4.19	6.310469
2292	CCCCCCOc1ccc(cc1)N(c1ccc(OCCCCCC)cc1)c1cc2c(cc1OCCCCC)n(CCCCCC)c1ccc(cc21)-c1cc2n(-c3cccc(OCCCCCC)c3)c3cc(C=C(C#N)C(O)=O)sc3c2s1	723	699.2501	13.5	12.74807	0.69	0.6741575	6.73	6.116626
2323	OC(=O)C(=C\c1ccc(s1)-c1ccc2n(-c3ccc(cc3)N(c3ccc(cc3)-n3c4ccccc4c4ccccc34)c3ccc(cc3)-n3c4ccccc4c4ccccc34)c3ccccc3c2c1)\C#N	735	655.1382	13.4	9.884835	0.66	0.7081243	6.53	4.508874
2325	OC(=O)CN1C(=S)S\C(=C/c2ccc(s2)-c2ccc3n(-c4ccc(cc4)N(c4ccc(cc4)-n4c5ccccc5c5ccccc45)c4ccc(cc4)-n4c5ccccc5c5ccccc45)c4ccccc4c3c2)C1=O	580	664.5068	9.2	10.255343	0.65	0.6822879	3.47	4.323113
2419	CCC1(CC)c2cc(ccc2-c2ccc(cc12)N(c1ccccc1)c1ccc2-c3ccc(cc3C(CC)(CC)c2c1)-c1ccc(\C=C(/C#N)C(O)=O)s1)N(c1ccccc1)c1ccc2-c3ccc(cc3C(CC)(CC)c2c1)-c1ccc(\C=C(/C#N)C(O)=O)s1	658	596.6453	12.7	9.569455	0.6	0.6439552	4.98	3.372473
2420	CCC1(CC)c2cc(ccc2-c2ccc(cc12)N(c1ccccc1)c1ccc2-c3ccc(cc3C(CC)(CC)c2c1)-c1ccc(s1)-c1ccc(\C=C(/C#N)C(O)=O)s1)N(c1ccccc1)c1ccc2-c3ccc(cc3C(CC)(CC)c2c1)-c1ccc(s1)-c1ccc(\C=C(/C#N)C(O)=O)s1	638	583.7627	13.3	9.249293	0.63	0.6263858	5.29	3.559488
2421	CCC1(CC)c2ccc(cc2-c2ccc(cc12)C(c1ccccc1)c1ccccc1)N(c1ccccc1)c1ccc2-c3ccc(cc3C(CC)(CC)c2c1)-c1ccc(\C=C(\C#N)C(O)=O)s1	678	635.68	10.1	10.938338	0.61	0.6503027	4.22	4.901772

Molecule ID	Smiles	V_{OC}	Previsão V_{OC}	J_{SC}	Previsão J_{SC}	FF	Previsão FF	PCE	Previsão PCE
2422	CCC1(CC)c2ccc(cc2-c2ccc(cc12)C(c1ccccc1)c1ccccc1)N(c1ccccc1)c1ccc2-c3ccc(cc3C(CC)(CC)c2c1)-c1ccc(s1)-c1ccc(\C=C(\C#N)C(O)=O)s1	646	632.0133	12.5	11.290004	0.62	0.6371421	4.97	4.479187
2433	CCCCC1(CCCC)c2cc(ccc2-c2ccc(cc12)-c1ccc(s1)-c1ccc(\C=C(/C#N)C(O)=O)s1)N(c1ccc2-c3ccc(cc3C(CCCC)(CCCC)c2c1)-c1ccc(s1)-c1ccc(\C=C(/C#N)C(O)=O)s1)c1ccc2n(CC)c3ccccc3c2c1	512	557.9976	3.72	6.876563	0.63	0.6491111	1.21	2.860853
2524	CCCCC1(CCCC)c2cc(ccc2-c2ccc(cc12)-c1ccc(\C=C(/C#N)C(O)=O)s1)N(c1ccc2-c3ccc(cc3C(CCCC)(CCCC)c2c1)-c1ccc(\C=C(/C#N)C(O)=O)s1)c1ccc2n(CC)c3ccccc3c2c1	497	570.0503	2.04	6.020715	0.68	0.6416252	0.69	2.457451
2600	CCCCn1c2ccc(\C=C(/C#N)C(O)=O)cc2c2nc3ccccc3nc12	518	533.9472	2.35	4.081291	0.63	0.6597917	0.77	1.34399
2601	CCCCn1c2ccc(\C=C(/C#N)C(O)=O)cc2c2nc3ccccc3nc12	500	538.6616	2.66	3.953623	0.64	0.6527568	0.86	1.328936
2602	CCCCn1c2ccc(cc2c2nc3ccccc3nc12)-c1ccc(\C=C(/C#N)C(O)=O)s1	602	568.5315	7.87	8.437211	0.65	0.6413458	3.07	3.050832
2603	CCCCn1c2ccc(cc2c2nc3ccccc3nc12)-c1ccc(\C=C(/C#N)C(O)=O)s1	579	580.9685	9.29	7.518244	0.64	0.6471708	3.45	2.763438
2604	CCCCn1c2ccc(cc2c2nc3ccccc3nc12)C#Cc1ccc(\C=C(/C#N)C(O)=O)cc1	454	567.0022	2.5	5.166845	0.56	0.6728543	0.63	1.887363
2707	OC(=O)C(=C\c1ccc(s1)-c1ccc(s1)-c1ccc(-c2ccc(cc2)N(c2ccccc2)c2ccccc2)c2nc(-c3ccccc3)c(nc12)-c1ccccc1)\C#N	690	638.4384	17.7	13.655347	0.66	0.6471506	8.06	5.578694
2708	CCc1ccc(cc1)N(c1ccc(CC)cc1)c1ccc(cc1)-c1ccc(-c2ccc(\C=C(/C#N)C(O)=O)s2)c2nc(-c3ccccc3)c(nc12)-c1ccccc1	670	654.5343	16.9	12.121122	0.66	0.6527077	7.47	5.432976
2709	OC(=O)C(=C\c1ccc(s1)-c1ccc(-c2ccc(s2)-c2ccc(s2)-c2ccc(cc2)N(c2ccccc2)c2ccccc2)c2nc(-c3ccccc3)c(nc12)-c1ccccc1)\C#N	620	641.3717	15.8	13.535375	0.64	0.6539216	6.27	5.872215

Molecule ID	Smiles	V_{OC}	Previsão V_{OC}	J_{SC}	Previsão J_{SC}	FF	Previsão FF	PCE	Previsão PCE
2710	CCC1(CC)c2ccccc2-c2ccc(cc12)N(c1ccc(cc1)-c1ccc(-c2ccc(\C=C(/C#N)C(O)=O)s2)c2nc(-c3ccccc3)c(nc12)-c1ccccc1)c1ccc2-c3ccccc3C(CC)(CC)c2c1	650	621.9894	15.6	11.055629	0.65	0.6521588	6.59	4.797169
2823	CCCCCCn1c2-c3ccccc3C(CCCCCC)(CCCCCC)c2c2cc(ccc12)-c1ccc(\C=C(/C#N)C(O)=O)cc1	813	730.849	11	11.275443	0.7	0.6827631	6.29	6.094091
2824	CCCCCCn1c2-c3ccccc3C(CCCCCC)(CCCCCC)c2c2cc(ccc12)-c1ccc(\C=C2/SC(=S)N(CC(O)=O)C2=O)cc1	700	680.4144	4.23	9.893745	0.75	0.6773288	2.22	4.578816
2825	CCCCCCn1c2-c3ccccc3C(CCCCCC)(CCCCCC)c2c2cc(ccc12)-c1ccc(\C=c2/sc([nH]c2=O)=C(C#N)C#N)cc1	695	714.7609	7.1	10.587411	0.73	0.6732989	3.6	5.545086
2826	CCCCCCn1c2-c3ccccc3C(CCCCCC)(CCCCCC)c2c2cc(ccc12)-c1ccc(\C=c2/sc([nH]c2=O)=C(C#N)C#N)s1	707	699.0174	11.9	10.718215	0.64	0.6924303	5.41	5.123736
2834	CCCCC(CC)Cn1c2cc(ccc2c2nc3c(ccc(-c4ccc(cc4)N(c4ccccc4)c4ccccc4)c3nc12)-c1ccc(cc1)N(c1ccccc1)c1ccccc1)-c1ccc(\C=C(/C#N)C(O)=O)o1	817	730.7593	12.9	10.376478	0.67	0.6754431	7.09	5.1437
2835	CCCCC(CC)Cn1c2cc(ccc2c2nc3c(ccc(-c4ccc(cc4)N(c4ccccc4)c4ccccc4)c3nc12)-c1ccc(cc1)N(c1ccccc1)c1ccccc1)-c1ccc(\C=c2/sc([nH]c2=O)=C(C#N)C#N)s1	707	691.0145	9.03	10.474021	0.68	0.6665969	4.35	4.761845
2836	CCCCC(CC)Cn1c2cc(ccc2c2nc3c(ccc(-c4ccc(cc4)N(c4ccccc4)c4ccccc4)c3nc12)-c1ccc(cc1)N(c1ccccc1)c1ccccc1)-c1ccc(\C=c2/sc([nH]c2=O)=C(C#N)C#N)o1	724	702.3038	9.71	10.820794	0.68	0.6731556	4.81	5.413929
2837	CCCCC(CC)Cn1c2cc(ccc2c2nc3c(ccc(-c4ccc(cc4)N(c4ccccc4)c4ccccc4)c3nc12)-c1ccc(cc1)N(c1ccccc1)c1ccccc1)-c1ccc(\C=C(/C#N)C(O)=O)s1	712	766.9654	10.8	10.654741	0.65	0.6665398	5.02	5.461783

Molecule ID	Smiles	V_{OC}	Previsão V_{OC}	J_{SC}	Previsão J_{SC}	FF	Previsão FF	PCE	Previsão PCE
2838	CCCCC(CC)Cn1c2cc(ccc2c2nc3c(ccc(-c4ccc(cc4)N(c4ccccc4)c4ccccc4)c3nc12)-c1ccc(cc1)N(c1ccccc1)c1ccccc1)-c1ccc(\C=C(/C#N)C(O)=O)s1	774	753.7779	10	10.889238	0.67	0.6618475	5.19	5.421282
2839	CCCCC(CC)Cn1c2cc(ccc2c2nc3c(ccc(-c4ccc(cc4)N(c4ccccc4)c4ccccc4)c3nc12)-c1ccc(cc1)N(c1ccccc1)c1ccccc1)-c1ccc(\C=C(/C#N)C(O)=O)s1	769	753.9289	10.3	10.7847	0.69	0.6578312	5.44	5.354717
2840	CCCCC(CC)Cn1c2cc(ccc2c2nc3c(ccc(-c4ccc(cc4)N(c4ccccc4)c4ccccc4)c3nc12)-c1ccc(cc1)N(c1ccccc1)c1ccccc1)-c1ccc(\C=C(/C#N)C(O)=O)s1	732	762.6386	10.7	10.735712	0.67	0.6621288	5.24	5.412281
2852	CCCCn1c2ccc(cc2c2nc3ccccc3nc12)C#Cc1ccc(\C=C(/C#N)C(O)=O)cc1	543	511.6313	4.43	4.146224	0.68	0.6053248	1.65	1.266589
2857	CCCCn1c2ccc(cc2c2nc3ccccc3nc12)C#Cc1ccc(s1)-c1ccc(\C=C(/C#N)C(O)=O)s1	543	572.153	6.77	8.021996	0.63	0.6362581	2.3	3.032843
2858	CCCCn1c2ccc(cc2c2nc3ccccc3nc12)C#Cc1ccc(s1)-c1ccc(\C=C(/C#N)C(O)=O)s1	562	559.3355	7.77	7.515769	0.62	0.6425162	2.68	2.784491
2902	CCCCC(CC)Cn1c2cc(ccc2c2nc3c(ccc(-c4ccc(cc4)N(c4ccccc4)c4ccccc4)c3nc12)-c1ccc(cc1)N(c1ccccc1)c1ccccc1)-c1ccc(\C=C(/C#N)C(O)=O)s1	797	745.3905	11.9	10.426394	0.64	0.6699523	6.05	5.210278
2948	CCCCCCc1cc(sc1C=C(C#N)C(O)=O)-c1sc(cc1CCCCCC)-c1cc2n(CC)c3ccccc3c2s1	700	690.6712	13.8	13.083149	0.77	0.677151	7.4	6.551175
2949	CCCCCCc1cc(sc1\C=C(/C#N)C(O)=O)-c1sc(cc1CCCCCC)-c1sc(cc1CCCCCC)-c1cc2n(CC)c3ccccc3c2s1	700	688.4568	14.6	14.254548	0.76	0.7127575	7.8	6.969052
2950	CCCCCCc1cc(\C=C(/C#N)C(O)=O)sc1-c1sc(cc1CCCCCC)-c1sc(cc1CCCCCC)-c1sc(cc1CCCCCC)-c1cc2n(CC)c3ccccc3c2s1	660	705.9118	15	13.859882	0.74	0.7220161	7.3	7.022703
3057	CCCCCCCCn1c2ccccc2c2c1c1c3ccccc3n(CCCCCCCC)c1c1c3cc(ccc3n(CCCCCCCC)c1c1c3ccccc3n(CCCCCCCC)c21)-c1ccc(\C=C(/C#N)C(O)=O)s1	750	739.3112	12.1	12.087546	0.64	0.6746369	5.79	6.174491

Molecule ID	Smiles	V_{OC}	Previsão V_{OC}	J_{SC}	Previsão J_{SC}	FF	Previsão FF	PCE	Previsão PCE
3059	CCCCCCCCn1c2ccccc2c2c1c1c3ccccc3n(CCCCCCCC)c1c1c3cc(ccc3n(CCCCCCCC)c1c1c3ccccc3n(CCCCCCCC)c21)-c1ccc(s1)-c1ccc(\C=C(/C#N)C(O)=O)s1	762	729.6123	13	12.26664	0.65	0.6675152	6.46	6.013843
4317	CCCC[C@H](CC)COc1ccc(cc1)-c1ccc2N([C@@H]3CC=C(C=C3)c3ccc(cc3)C(O)=O)c3ccc(cc3Sc2c1)-c1ccc(OC[C@@H](CC)CCCC)cc1	627	696.1801	1.49	7.696134	0.78	0.725555	0.73	3.955122
4318	CCCC[C@H](CC)COc1ccc(cc1)-c1ccc2N(c3ccc(cc3)C(O)=O)c3ccc(cc3Sc2c1)-c1ccc(OC[C@@H](CC)CCCC)cc1	671	696.3722	1.93	6.282032	0.76	0.7452404	0.99	3.266078
4325	CCCCCCCCCCCCN1c2ccc(\C=C\c3ccc(cc3)N(c3ccccc3)c3ccccc3)cc2Oc2cc(\C=C(/C#N)C(O)=O)ccc12	733	722.8973	14.7	11.872721	0.71	0.7237197	7.7	6.01699
4326	CCCCCCCCCCCCN1c2ccccc2Oc2cc(\C=C(/C#N)C(O)=O)ccc12	722	703.5102	11.5	8.589032	0.75	0.7108254	6.2	4.569663
4343	CCCCCCn1c2ccccc2c2cc(ccc12)N1c2ccccc2Sc2cc(\C=C3/SC(=S)N(CC(O)=O)C3=O)ccc12	642	665.9558	11.45	9.944632	0.72	0.6817518	5.26	4.255093
4344	CCCCn1c2ccccc2c2cc(ccc12)N1c2ccccc2Sc2cc(\C=C3/SC(=S)N(CC(O)=O)C3=O)ccc12	656	617.4946	10.01	10.584902	0.71	0.6905141	4.67	4.473421
4345	CCCCCCCCn1c2ccccc2c2cc(ccc12)N1c2ccccc2Sc2cc(\C=C(/C#N)C(O)=O)ccc12	723	719.7197	14.43	13.900985	0.68	0.7013939	7.09	6.533822
4346	CCCCn1c2ccccc2c2cc(ccc12)N1c2ccccc2Sc2cc(\C=C(/C#N)C(O)=O)ccc12	735	663.3631	16.45	12.676894	0.68	0.6902631	8.08	5.827111
4347	CCn1c2ccccc2c2cc(ccc12)N1c2ccccc2Sc2cc(\C=C(/C#N)C(O)=O)ccc12	735	634.0545	15.06	12.13092	0.68	0.6828222	7.47	4.811954
4358	CCCCCCn1c2-c3ccccc3C(CCCCCC)(CCCCCC)c2c2cc(ccc12)-c1ccc(\C=C(/C#N)C(O)=O)o1	733	751.2059	13.7	13.360945	0.65	0.6799189	6.52	6.418753
4359	CCCCCCn1c2-c3ccccc3C(CCCCCC)(CCCCCC)c2c2cc(ccc12)-c1ccc(\C=C(/C#N)C(O)=O)s1	742	728.4987	14.6	12.819298	0.64	0.6854608	6.95	6.265274

Molecule ID	Smiles	V_{OC}	Previsão V_{OC}	J_{SC}	Previsão J_{SC}	FF	Previsão FF	PCE	Previsão PCE
4360	CCCCCCn1c2-c3ccccc3C(CC)(CC)c2c2cc(ccc12)-c1ccc(\C=C(/C#N)C(O)=O)o1	763	702.8867	15.8	12.38066	0.63	0.6785513	7.64	5.998233
4361	CCCCCCn1c2-c3ccccc3C(CC)(CC)c2c2cc(ccc12)-c1ccc(\C=C(/C#N)C(O)=O)s1	710	699.2221	15.6	13.142677	0.67	0.6614239	7.39	6.160211

Tabela 6.1 Tabela de dados do conjunto de treino do modelo Ts1

6.2 **Previsões do conjunto de teste para o modelo Ts1**

Molecule ID	Smiles	V_{OC}	Previsão V_{OC}	J_{SC}	Previsão J_{SC}	FF	Previsão FF	PCE	Previsão PCE
71	CC1(C)CCN2CCC(C)(C)c3c2c1cc1cc(\C=C\c2ccc(\C=C(/C#N)C(O)=O)s2)c(=O)oc31	580	630.5489	14.89	11.278739	0.73	0.7290593	6.3	5.156849
72	CC1(C)CCN2CCC(C)(C)c3c2c1cc1cc(\C=C\c2ccc(\C=C(/C#N)C(O)=O)s2)c(=O)oc31	630	630.5489	13.37	11.278739	0.77	0.7290593	6.5	5.156849
184	CCCCOc1ccc(c(OCCCC)c1)-c1ccc2N(CCCC)c3ccc(cc3Sc2c1)-c1ccc(s1)-c1ccc(\C=C(/C#N)C(O)=O)s1	706	655.0087	16.36	12.776116	0.65	0.6674268	7.5	6.054969
239	CCCCCCOc1ccc(c(OCCCCCC)c1)-c1ccc2N(CCCCCC)c3ccc(cc3Sc2c1)-c1sc(cc1CCCCCC)-c1cc[n+](cc1)C(O)=O	715	697.2339	12.4	8.968811	0.74	0.6725313	6.6	4.568646
241	CCCCCCOc1ccc(c(OCCCCCC)c1)-c1ccc2N(CCCCCC)c3ccc(cc3Sc2c1)-c1sc(cc1CCCCCC)-c1cc[n+](cc1)C(O)=O	706	697.2339	13.1	8.968811	0.72	0.6725313	6.7	4.568646
494	CCCCN1c2ccc(C=C3SC(=S)N(CC(O)=O)C3=O)cc2Sc2cc(ccc12)-n1c2ccc(cc2c2cc(ccc12)C(C)(C)C)C(C)(C)C	648	647.3164	10.03	9.120871	0.66	0.649077	4.38	3.854017
766	CCCCOc1ccc(cc1)N(c1ccc(OCCCC)cc1)c1ccc2c(c1)n(CCCC)c1cc(ccc21)-c1ccc(\C=C(/C#N)C(O)=O)s1	520	667.1081	4.47	13.041481	0.58	0.6492725	1.36	5.750339

Molecule ID	Smiles	V_{OC}	Previsão V_{OC}	J_{SC}	Previsão J_{SC}	FF	Previsão FF	PCE	Previsão PCE
771	CCCCCCCCOc1ccc(cc1)-c1c2n(CCCCCCCC)c3ccc(cc3c2c(-c2ccc(OCCCCCCCC)cc2)c2n(CCCCCCCC)c3ccccc3c12)-c1ccc(\C=C(/C#N)C(O)=O)s1	757	732.0587	14.1	13.53086	0.71	0.6791715	7.58	6.550976
1088	CCCCn1c(nc2c(sc(-c3cccs3)c12)-c1cccs1)-c1ccc(cc1)-c1ccc(\C=C(/C#N)C(O)=O)s1	575	579.6705	5.75	8.991203	0.66	0.6476887	2.19	3.672466
1394	COc1ccc(cc1)N(c1ccc(OC)cc1)c1ccc(\C=C\c2nc3ccc(cc3nc2\C=C\c2ccc(cc2)N(c2ccc(OC)cc2)c2ccc(OC)cc2)C(O)=O)cc1	650	647.878	8.72	9.313555	0.73	0.6748946	4.1	4.271873
1833	CCCCCCOc1ccc(cc1)N(c1ccc(OCCCCCC)cc1)c1ccc(cc1)-c1sc(cc1CCCCCC)-c1cc[n+](CC(O)=O)cc1	670	695.5114	11.9	9.329562	0.74	0.6841933	5.9	4.582058
1836	CCCCCCOc1ccc(cc1)N(c1ccc(OCCCCCC)cc1)c1ccc(cc1)-c1sc(cc1CCCCCC)-c1cc[n+](CC(O)=O)cc1	697	695.5114	11.3	9.329562	0.74	0.6841933	5.8	4.582058
1941	CCCCCCc1cc(sc1-c1ccc(cc1)N(c1ccccc1)c1ccccc1)-c1cc[n+](CC(O)=O)cc1	643	687.9045	6.6	7.353908	0.74	0.6668424	3.1	3.490195
1944	CCCCCCc1cc(sc1-c1ccc(cc1)N(c1ccccc1)c1ccccc1)-c1cc[n+](CC(O)=O)cc1	673	687.9045	7.2	7.353908	0.75	0.6668424	3.6	3.490195
1974	CCCCCCOc1ccc(cc1)N(c1ccc(cc1)-c1cc2n(-c3ccccc3OCCCCCC)c3cc(\C=C(/C#N)C(O)=O)sc3c2s1)c1cc2c(cc1OCCCCCC)n(CCCCCC)c1ccccc21	754	702.9049	15	13.568442	0.69	0.6800729	7.8	6.26306
2297	CCCCCCOc1cccc(c1)-n1c2ccsc2c2sc(cc12)-c1ccc(cc1)N(c1ccc(cc1)-c1cc2n(-c3cccc(OCCCCCC)c3)c3ccsc3c2s1)c1cc2c(cc1OCCCCCC)n(CCCCCC)c1ccc(cc21)-c1cc2n(-c3cccc(OCCCCCC)c3)c3cc(C=C(C#N)C(O)=O)sc3c2s1	709	669.8815	13.2	11.885453	0.67	0.6625574	6.27	5.329988

Molecule ID	Smiles	V_{OC}	Previsão V_{OC}	J_{SC}	Previsão J_{SC}	FF	Previsão FF	PCE	Previsão PCE
2853	CCCCn1c2ccc(cc2c2nc3ccccc3nc12)C#Cc1cccc(\C=C(/C#N)C(O)=O)c1	594	542.2235	2.41	5.329967	0.68	0.6440083	0.97	1.906962
2854	CCCCn1c2ccc(cc2c2nc3ccccc3nc12)C#Cc1cccc(\C=C(/C#N)C(O)=O)c1	593	542.2235	2.73	5.329967	0.67	0.6440083	1.08	1.906962
2855	CCCCn1c2ccc(cc2c2nc3ccccc3nc12)C#Cc1ccc(\C=C(/C#N)C(O)=O)s1	566	561.024	6.52	7.750337	0.65	0.6483821	2.41	2.929115
2856	CCCCn1c2ccc(cc2c2nc3ccccc3nc12)C#Cc1ccc(\C=C(/C#N)C(O)=O)s1	568	561.024	7.38	7.750337	0.65	0.6483821	2.72	2.929115
4342	CCCCCCCCn1c2ccccc2c2cc(ccc12)N1c2ccccc2Sc2cc(\C=C3/SC(=S)N(CC(O)=O)C3=O)ccc12	642	661.8365	11.26	10.643579	0.68	0.6986299	4.91	4.718475

Tabela 6.2 Tabela de dados do conjunto de teste do modelo Ts1